

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 8532.25 Seconds

(without alignments)
10921.753 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202
Perfect score: 3202
Sequence: 1 atgtgggtatctatctgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBml:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	3134.2	97.9	84196	8	ATT3A5
2	1132	35.4	4818	8	AF044216
3	176.8	5.5	191	11	AL844732
4	133.2	4.2	104992	2	AC005504
5	133.2	4.2	169546	2	AC004157
6	130.8	4.1	104992	2	AC005504
7	130.8	4.1	169546	2	AC004157
8	129.4	4.0	14867	2	AE001398
9	128.6	4.0	172816	9	AC093899
10	127.8	4.0	86827	3	PFMAL3P5
11	126.4	3.9	133877	2	AC120883
12	125.4	3.9	86827	3	PFMAL3P5
13	124.6	3.9	205130	2	AC105425
14	124.4	3.9	4601	3	DMU11584
15	124.4	3.9	19517	3	DMU37541
16	123.4	3.9	67970	3	PFMAL1P3
17	122.6	3.8	4601	3	DMU11584
18	122.6	3.8	19517	3	DMU37541
19	121.2	3.8	177293	2	AC069525
20	121	3.8	172816	9	AC093899
21	121	3.8	176970	3	AC104073
22	118	3.7	84196	8	ATT3A5
23	118	3.7	72243	9	AL731858
24	116.4	3.6	205130	2	AC105425
25	115.6	3.6	133877	2	AC120883
26	115.2	3.6	176970	3	AC104073
27	114.4	3.5	14867	3	AE001398
28	113	3.5	137889	9	AC073269
29	112.6	3.5	206059	2	AC127383
30	112.2	3.5	104853	9	AC117444
31	111.6	3.5	111861	9	AC069435
32	111.4	3.4	15656	3	AB070264
33	110.2	3.4	15656	3	AB070264
34	109.6	3.4	159618	9	AC096750
35	107.8	3.4	152209	9	HS1108D11
36	107	3.3	105238	9	AC011458
37	106.6	3.3	106958	9	AL807813
38	106.4	3.3	155106	9	AC104069
39	106.2	3.3	181792	9	AC098822
40	105.6	3.3	176174	9	AC007483
41	105.6	3.3	177293	2	AC069525
42	105.6	3.3	206059	2	AC127383
43	105.2	3.3	114276	9	AC011355
44	104.8	3.3	127439	9	AC126178
45	104.6	3.3	127439	9	AC126178

ALIGNMENTS

RESULT 1
LOCUS ATT3A5/c
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
ACCESSION AL132979
VERSION AL132979.2
KEYWORDS GI:6782244
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 84196)
Bloeker, H., Mewes, H.W., Lemcke, R., Mayer, K.F.X., Quetler, F. and
Salanoubat, M.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84196)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemc@emaps.biochem.mpg.de, mayer@emaps.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr

COMMENT On Jan 27, 2000 this sequence version replaced gi:6434247. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
source
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983. .3338
/note="355 bp LTR"
3339. .7564
/note="LTR-transposon"
7565. .7922
/note="367bp LTR"
12339. .13197
/gene="T3A5.10"
join(12339. .12665,12742. .12803,12888. .13055,13125. .13197)
/gene="T3A5.10"
/note="similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AT094772_1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAAVRRREDDVYBENCVTITYKRRKMEBEVDLVESRITLPCV QATNRGIVARNASGASSETSVIVRRRDSPEVEQCIIEEDSSVSCCSTSEKSKRR IEFVLDENNDDRETETSWIYDNLKSESNMDSVAVEDVESRRRLKSLHETV KEAELDFEQVAKDLRNLKLECSMKYNFDEKDEPLGGGYEMVKLPN"
12339. .12665
/gene="T3A5.10"
/number=1
12666. .12741
/gene="T3A5.10"
/number=1
12742. .12803
/gene="T3A5.10"
/number=2
12804. .12887
/gene="T3A5.10"
/number=2
12888. .13055
/gene="T3A5.10"
/number=3
13056. .13124
/gene="T3A5.10"
/number=3
13125. .13197
/gene="T3A5.10"
/number=4
13198. .16470
/gene="T3A5.20"
complement(join(15858. .15901,16014. .16470))
/note="similarity to predicted proteins, Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEMHNSMGCAARSFSEFVEIRSEKSNKSNINRSRSSF MSADAFSVTDISMLPADEIFLKGKILPKETSHVHRTLGEEELLTEEGSMVDGNTF

SLRPILSSSSFSSTKGTWRELLGKRTHVSKKTDKYNEEVLSDDKHITSGNVAIIRBC QVADTR"
complement(15858. .15901)
/number=1
complement(15902. .16013)
/number=1
complement(16014. .16470)
/number=2
18303. .19931
/gene="T3A5.30"
complement(18303. .19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAVMCTDSGLNMAIAOOLIKOKOOOSOHOOEOEORPEPPNPN PSFGFTLPGSGFSDPQVNTNDPGFHHVHONAAVASSEPPDSEMEELINGDAS QTPNDPRITGHDPFVSFPRSLSPSYINRNKDDASQOLPPPPASIAIYSPSPSPQ HPPPPPPDPDNLQPTFKAIHDYARKPEPTDPLIRIKSVESGDIORGVITFME ALSHKETESPSSSSSLEDFTLSTKTLNACYSKFAHLTAQAALIEATNOSNTHI VDEGIFGQIOWSALQALATRSSGKPTRISGIPAPSLDSGSPSLIATGNRLRDA AILDNFEFYPVLTPIOLNNGSSFRVDQDVLVNFMLELYKLDELATVGTALRLA RSLNPRIVLTIGEYVSLNRYEPANRYKNSLRFSVAFESLEPDLDRSKRLRYEVL FGRRINDLVSDDDNNKPGTFRGLMEKEKEDMRILMEKAGEPYKPSRYAVSOAKLLIM NINISTLVSESPGFTSLAMNINPFLITVSMR"
complement(18303. .19931)
/number=1
26093. .28999
/gene="T3A5.40"
complement(join(26093. .26338,27091. .27200,27308. .27386, 27478. .27570,27723. .28037,28136. .28288,28374. .28698, 28779. .28999))
/note="Contains Cytochrome P450 cysteine heme-1iron ligand signature AA45-464;Prokaryotic membrane lipoprotein lipid attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEETHEHTLPLPLPSLSLLEFLILKRRNKTRFNPFGKS GMPFLGFTIGYLRKYATTTGDDPMQOHSYKGIKRSNLGEPPIVASADGLNRFTIQ NEGRLFECSPRSIGIGITGKSMVLVGDHMRMRSTSLNPLSHARLRTILLDVERH TLFVLDSMOONSIFSADDEAKKFTFNMAKHISMDSGDEETSQLKEYTTFKGYVS APJNLPEIATYHKAQSRAITLTKTERMEERKLIDKEDEEVEVKEDAEBSKSDH VRKQRDIDDLLGWLVKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFFLQACPAV EELREHELEIYARAKKELGESELMWDYKKKDFQCVINETLRGNNVRFJHRKALDY RYKGYDIPSGMKVLPVISAVALHDSRYDQNLNPNPMQOONNGASSGSGSFTWGN NVMPPGGGRPLCAGSELAKLEMAVFTHHLVLKRNWLEADDKCFAPFVDFPGLPIR VSRIL"
complement(26093. .26338)
/number=1
complement(26339. .27090)
/number=1
complement(27091. .27200)
/number=2
complement(27201. .27307)
/number=2
complement(27308. .27386)
/number=3
complement(27387. .27477)
/number=3
complement(27478. .27570)
/number=4
complement(27571. .27722)
/number=4
complement(27723. .28037)
/number=5
complement(28038. .28135)
/number=5
complement(28136. .28288)
/number=6
complement(28289. .28373)
/number=6

QY	421	AGTTTATTGATCATATATACATGATCATTTCTTGATTATCTATGAAGATTTCTTAC	480
Db	31780	AGGTTATTTTGATCATATATATACATGATCATTTCTTGATTATCTATGAAGATTTCTTAC	31721
QY	481	CAATTAAATTTGGAAATTCATATCTCTGTATTTAAATTAATTAATACGATGGAATATAC	540
Db	31720	CAATTAAATTTGGAAATTCATATCTCTGTATTTAAATTAATTAATACGATGGAATATAC	31661
QY	541	GTTATATCGATCTCCATCATATGATATGATTTCTTGCTTAATCCAGCAATATTATTAACA	600
Db	31660	GTTATATCGATCTCCATCATATGATATGATTTCTTGCTTAATCCAGCAATATTATTAACA	31601
QY	601	AGAGTATTGAAAAAAACCGAAAAATAGAAAAAGGAAAGATATGATCCATGAGTATG	660
Db	31600	AGAGTATTGAAAAAAACCGAAAAATAGAAAAAGGAAAGATATGATCCATGAGTATG	31541
QY	661	TGATTAATTTATCAAAAGAAATAGAGATATGACAAACCAAAAGTTGGATTAATATGCTCCT	720
Db	31540	TGATTAATTTATCAAAAGAAATAGAGATATGACAAACCAAAAGTTGGATTAATATGCTCCT	31481
QY	721	GCCAGCTTTTCTCACATCATATATGACCCCTTATTTGGATTTCTGGATATATCGTTAAAA	780
Db	31480	GCCAGCTTTTCTCACATCATATATGACCCCTTATTTGGATTTCTGGATATATCGTTAAAA	31421
QY	781	TTTTCGATTAACGATTTGTGAAAAAATATTTTATTTTGTAGCTGATCTCAATATTTATGTCCA	840
Db	31420	TTTTCGATTAACGATTTGTGAAAAAATATTTTATTTTGTAGCTGATCTCAATATTTATGTCCA	31361
QY	841	GGTATTTTGCAATATCTCTGTTTAAAGCATATTTGTCTTCTTTTGTGCTGTTCTCT	900
Db	31360	GGTATTTTGCAATATCTCTGTTTAAAGCATATTTGTCTTCTTTTGTGCTGTTCTCT	31301
QY	901	TAACTATATATATATCGGGATATATGATTAATCAATGATATATCAACAAACAAATATGCTGGG	960
Db	31300	TAACTATATATATATCGGGATATATGATTAATCAATGATATATCAACAAACAAATATGCTGGG	31241
QY	961	ACCAATTTTGAATAAACTTTTCTCAAAACATTTACGGGACATGAGCTCGACCTTAAAAATA	1020
Db	31240	ACCAATTTTGAATAAACTTTTCTCAAAACATTTACGGGACATGAGCTCGACCTTAAAAATA	31182
QY	1021	CGATTTTACAGCGCTACTACTAGATTTACTACATTAAGCATTAAGAGCCCGTTCAAGC	1080
Db	31181	CGATTTTACAGCGCTACTACTAGATTTACTACATTAAGCATTAAGAGCCCGTTCAAGC	31122
QY	1081	TATTTATTAACAAGTTACAAACTGAAATTTACTCTTGAATCTCTTAGAAAAATTTTGGAAATA	1140
Db	31121	TATTTATTAACAAGTTACAAACTGAAATTTACTCTTGAATCTCTTAGAAAAATTTTGGAAATA	31062
QY	1141	CCGGTGTATATGAATATATAGATTTACTGTGTAACAAATATGTTAATCAATTAAGTGCTCA	1200
Db	31061	CCGGTGTATATGAATATATAGATTTACTGTGTAACAAATATGTTAATCAATTAAGTGCTCA	31002
QY	1201	ACATATACATATATTCCTTACAGAAAAAAACAACTTAAAGAGATTAACATATCATATAT	1260
Db	31001	ACATATACATATATTCCTTACAGAAAAAAACAACTTAAAGAGATTAACATATCATATAT	30942
QY	1261	GGGATATGCTATACCTTACAGTATGCTATCTAGAGACTTAAGAAATATGTTATGTATGTTC	1320
Db	30941	GGGATATGCTATACCTTACAGTATGCTATCTAGAGACTTAAGAAATATGTTATGTATGTTC	30882
QY	1321	GATAAATGAAATTCACACGCGTGTATTAATTAATTTGGAGCCGATATGTTACGATCACTGCAA	1380
Db	30881	GATAAATGAAATTCACACGCGTGTATTAATTAATTTGGAGCCGATATGTTACGATCACTGCAA	30822
QY	1381	ATATCATTTCTTGTTGGTCAACAATTAATAAACAAACAGAAAAAGAAACGATTTT	1440
Db	30821	ATATCATTTCTTGTTGGTCAACAATTAATAAACAAACAGAAAAAGAAACGATTTT	30762
QY	1441	CTGGATTCACATTAATGATCTTAATTAATGATATGATCTTTGGTTTCAATTCGAAATGCC	1500
Db	30761	CTGGATTCACATTAATGATCTTAATTAATGATATGATCTTTGGTTTCAATTCGAAATGCC	30702

[illegible]

[illegible]

QY	2551	ACCCCTACTCGT	TAGAGCCCA	ACCCCTTTTCCCA	TATCCCTAAGG	TAATTTTGGAAATCCCAAT	2610
Db	481	ACCCCTACTCGT <th>GTAGCCCA</th> <th>ACCCCTTTTCCCA</th> <th>TATCCCTAAGG</th> <th>TAATTTTGGAAATCCCAAT</th> <th>540</th>	GTAGCCCA	ACCCCTTTTCCCA	TATCCCTAAGG	TAATTTTGGAAATCCCAAT	540
OY	2611	TTAAACGATTTGA <td>ACCGGTAC</td> <td>CGGAGCTTCCG</td> <td>GGAATTCGCTG</td> <td>AGCATTTATCAAAAT</td> <td>2670</td>	ACCGGTAC	CGGAGCTTCCG	GGAATTCGCTG	AGCATTTATCAAAAT	2670
Db	541	TTAAACGATTTGA <td>ACCGGTAC</td> <td>CGGAGCTTCCG</td> <td>GGAATTCGCTG</td> <td>AGCATTTATCAAAAT</td> <td>600</td>	ACCGGTAC	CGGAGCTTCCG	GGAATTCGCTG	AGCATTTATCAAAAT	600
OY	2671	TATTAGCAGCAAT	GGGTTTAT	TAAATTTAAAAA	CTCACAACTTGAT	CAATAAAAATTCAT	2730
Db	601	TATTAGCAGCAAT	GGGTTTAT	TAAATTTAAAAA	CTCACAACTTGAT	CAATAAAAATTCAT	660
OY	2731	AAACACTTTTAC	GTGATTCGT	ACGATCTATCT	TAAGACTTTTTTTT	TTTCTACACGGT	2790
Db	661	AAACACTTTTAC	GTGATTCGT	ACGATCTATCT	TAAGACTTTTTTTT	TTTCTACACGGT	720
OY	2791	GGATGAAGATTA	TAGTACTATT	TACCCAGACAA	TTTGATTTATATCA	TTAATTC	2850
Db	721	GGATGAAGATTA	TAGTACTATT	TACCCAGACAA	TTTGATTTATATCA	TTAATTC	780
OY	2851	ATGATATTTAT	GATATTAAT	AGCTTTAACT	ATTTCAGATCCAG	CTTCTGCAACT	2910
Db	781	ATGATATTTAT	GATATTAAT	AGCTTTAACT	ATTTCAGATCCAG	CTTCTGCAACT	840
OY	2911	TTGTTTTTAAT	TATTAAGACT	TTTAATTAAT	TAAGATTAAGGAC	CATACGAGCAACA	2970
Db	841	TTGTTTTTAAT	TATTAAGACT	TTTAATTAAT	TAAGATTAAGGAC	CATACGAGCAACA	900
OY	2971	AAAGTAATGA	CACGAGGAA	CAAAAGCCAT	GAAAGCTATG	GTATTAAGCTTAATA	3030
Db	901	AAAGTAATGA	CACGAGGAA	CAAAAGCCAT	GAAAGCTATG	GTATTAAGCTTAATA	960
OY	3031	AGAGATTTTAT	TAAATTTTAA	TATGACAT	GTATTAACAAAT	TATTTTTCGACTTCTTTAAA	3090
Db	961	AGAGATTTTAT	TAAATTTTAA	TATGACAT	GTATTAACAAAT	TATTTTTCGACTTCTTTAAA	1020
OY	3091	ACCCCTCTT	TCAAAACAG	ACAGCTCCCTTTT	CAGTAGACAG	ATCCGATCTCTTAAAG	3155
Db	1021	ACCCCTCTT	TCAAAACAG	ACAGCTCCCTTTT	CAGTAGACAG	ATCCGATCTCTTAAAG	1080
OY	3151	ACAAAGCATTA	GAAGAAAG	GAAGTAGAG	GAGAGAGAG	GAGAGAAACTACTCTCC	3202
Db	1081	ACAAAGCATTA	GAAGAAAG	GAAGTAGAG	GAGAGAGAG	GAGAGAAACTACTCTCC	1132
RESULT 3	AL844732	191 bp	DNA	linear	STS 25-JUL-2002		
LOCUS	AL844732	Arabidopsis thaliana	transposon	insertion	STS SM.3.29533, sequence	tagged site.	
ACCESSION	AL844732						
VERSION	AL844732.1	GI:21998461					
KEYWORDS	STS: STS, sequence	tagged site.					
SOURCE	Arabidopsis thaliana.						
ORGANISM	Arabidopsis thaliana						
REFERENCE	1	Clarke, J.H., Bowles, B., Carter, J., Hart, D., McCullagh, B., Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.					
AUTHORS	2	(bases 1 to 191)					
JOURNAL	3	Submitted (25-JUL-2002)					
REFERENCE	4	Clarke, J.H., Bowles, B., Carter, J., Hart, D., McCullagh, B., Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.					
AUTHORS	5	Submitted (25-JUL-2002)					
TITLE	6	Submitted (25-JUL-2002)					
JOURNAL	7	Submitted (25-JUL-2002)					
COMMENT	8	Submitted (25-JUL-2002)					

[illegible]

[illegible]

	RESULT 7						
	AC004157/c						
LOCUS	AC004157	169546 bp	DNA	linear	HTG-12-AUG-2000		
DEFINITION	<i>Plasmodium falciparum</i> chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.						
ACCESSION	AC004157						
VERSION	AC004157.8	GI:9797712					
KEYWORDS	HTG; HTGS_PHASEI;						
SOURCE	<i>Plasmodium falciparum</i> .						
ORGANISM	<i>Plasmodium falciparum</i> .						
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 169546)						
AUTHORS	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T., Kudr,O.B., Conway,A.B. and Davis,R.W.						
TITLE	<i>Plasmodium falciparum</i> 3D7 chromosome 12						

[illegible]

OY	699	AGCTGTGGAAATGATGTCCTCCGACGCTTCTCTCAATCATATGACCCCTATTGG	758
Db	85724	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85665
OY	759	ATTTCTGATATTCGTTAAATTTGGATACGATGTGGAATAATTTTATTTGGTAG	818
Db	85664	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85623
OY	819	CTGATCTCAATATATGTCCTCCGATTTCCATATCTTCTGTTAAAGCATATTTGTC	878
Db	85622	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85564
OY	879	TTTCTTTTGTTCGTTCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT	938
Db	85563	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85504
OY	939	TATCAACAATTAATTTCTGCGACCATTTTGAATTAATTTCTCAACATTTGCGAC	998
Db	85503	ATATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85444
OY	999	ACTGACCTGACCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1058
Db	85443	ATATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85384
OY	1059	AGCATTAAGGACCCGTTCAACCTTTTATCAAAAGTCAACATTAATTAATTAATTAAT	1118
Db	85383	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85324
OY	1119	CTTTAGAAATTTTGGATTTACCGGTTTATGTAATTAATTAATTAATTAATTAATTAAT	1178
Db	85323	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85264
OY	1179	TA-----TGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1234
Db	85263	ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1294
OY	1235	TAAAGAAAGTTAATCAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1348
Db	85203	TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1408
OY	1295	AGACTAAGAA-----TAGTATGATGATGATTAATTAATTAATTAATTAATTAATTAAT	1468
Db	85143	ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1528
OY	1349	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1588
Db	85083	TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1648
OY	1409	AACAAACACAGAAAAAGAAAGATTTTCTGATGATTCATTAATTAATTAATTAATTAAT	1708
Db	85023	ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1768
OY	1469	CATGATCTTTGGGTTACAGTTTCGAGTCCTCAAGCGTTGATTAATTAATTAATTAATTAAT	1828
Db	84963	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1888
OY	1529	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1948
Db	84903	ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2008
OY	1589	GCGAATCTGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2068
Db	84843	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2128
OY	1649	TACTTGGATGTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2188
Db	84783	TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2248
OY	1709	GTAATAAAGTTTGTCTTATTTTCCGATGATTAATTAATTAATTAATTAATTAATTAAT	2308
Db	84723	ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2368
OY	1769	TTTGAATAATGTAACCTTTTACATTAATTAATTAATTAATTAATTAATTAATTAAT	2428
Db	1769	TTTGAATAATGTAACCTTTTACATTAATTAATTAATTAATTAATTAATTAATTAAT	2488

FEATURES

source

gene

CDS

/gene="PFB0490c"

/gene="PFB0490c"

/gene="PFB0490c"

/gene="PFB0490c"

/gene="PFB0490c"

OY	1455	AATGATCTAAAGGCATGAGATCTTTGGGTACGTTCCGAAGCCTCTCAAGCGGTGA	1514
Dd	24312	TATATAATATATTATATATATTAATACAGTATATATAAATATATTTATTTTACAGTTATA	2437
OY	1515	ACCATCGCAACTATTTAAATTCGTTCTTTAATGCAATCTTTAACATATTTATTTGGTAGTT	1574
Dd	24372	TAAATATATTTATATTTTATACAGTTATATATA-ATATATTTAAATTTTATACAGTTATAT	2443
OY	1575	GGAATTTAATAGACGCCAAGCTTGAAACATTACAATTTTAAATTAATTAATCTAGTAGTGA	1634
Dd	24431	AAATATATTTTAAATTTTAAACAGTTATATAAATAATTTATATTTTATACAGTTATATA	2449
OY	1635	TATTCCAAATACATCTTGGATGGTTTTAACTTAATCTTGTTCTCTCACGGATATAA	1694
Dd	24491	ATA-TATTAATATTTTATACAGTTATATAAATAATTTATGTTTTATACATTTATATATA	24549
OY	1695	TATTTATCATCCAGGTAATAAAAAAGTTTGCTTTATTTTGGCAGTCAGCAAGAGATAAAC	1754
Dd	24550	TATATTTATTTTATTTTATACATTTGGTTTAAATATATTTT---ATATATAAATATATTTTA	24606
OY	1755	TATGACCTTAAATTTTGTGAATAAGTAACCCCTTTACATCATGATTAATACCGTAGTT	1814
Dd	24607	TATTTAAATATTTTATATATAATATATATTTGATTTTAAATAATATTAATATATAATA	24666
OY	1815	TTTGTTGCCATTAAGCAAGCCTCTACAACTGTGATAGCAATTTTCTGCAAAATATTA	1874
Dd	24667	TATATGATTTTATTAATATGTTTAAATAAATAATATATTTTATTTAATATATATGTTTA	24726
OY	1875	ATTAGAATTCATGCTACTATATCAATAGAGAAGAAACAGCTGATTTACATTTTAATTTAA	1934
Dd	24727	TATATCAATATATATTTATATTTAATATATATTTTACATATTAATAATATTTAAATTTGA	24786
OY	1935	AGACAAAATTTTGCAAAATGTTTAAATTTCTAACAAATATTTAAATATGATGCGCTAT	1994
Dd	24787	TATATATTTTATATTTGATATATATTTTATATATTAATTAATATATTTACATTTGATATAT	24846
OY	1995	AATGATATTCCTATGTTCTTAAATATTTTTTTATATTTAGTATTAATACATTTGA	2054
Dd	24847	TTTATATATATTAATATATTTTACATTTGATATATTTTATATAATATTAATAATTTACA	24906
OY	2055	ACCAATA-ATAGTGGGAATTCAAATATCTCCATTAATTTTGTGAATCTACAATTT	2113
Dd	24907	TTTGATATATATTTTATATAATATTAATAATTAATTAATTTGATATATGTTTATAGATAAAT	24966
OY	2114	ATTAAATATTTAGTCAATTAACAATGATAGAAAAGTTCCAAAAAAATTTGTTAACGAAA	2173
Dd	24967	ATATATTTTCGTTTATATATATTTATTAATATAATATATTTACGTTTAAATATATTTT	25026
OY	2174	CTTCCAAATTTTTTTTTTTTATGGAACAAGAAATAACAGATAGAAAACATTTTGTGTG	2233
Dd	25027	ATCTATAATATATATTTTATGTTTATATATATTTATATATAATTAATATTTACGTTTA	25086
OY	2234	GAATGGAAGTACTAATATATCATTTAAGCAAAATTTTAAAAAAT	2274
Dd	25087	ATATATATTTTATAGATAATATATATTTACGTTTAAATAT	25127
RESULT 10			
PFWAL3P5			
LOCUS	Plasmodium falciiparum MAU3P5, complete sequence.	86627 bp DNA linear	INV 04-MAY-2000
DEFINITION	AI034556 AL008971 AL008972 AL008978 AL010141 AL010155 AL010162		
ACCESSION	AI010206 AL010210 AL139179		
VERSION	AL034556.3 GI:7711064		
KEYWORDS	HTE; centromere; CTRP protein; initiation factor E4; Serine/threonine protein phosphatase.		
SOURCE	Plasmodium falciiparum 3D7.		
ORGANISM	Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmidium.		
REFERENCE	1. (bases 1 to 86627) Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Church, S., Davies, A., De Haven, J., Drenth, J., Evans, R., Fraser, L., Gaitanaris, G., Gilbert, W., Hegerl, R., Johnson, D., Jones, K., Jones, M., Jones, N., Jones, P., Jones, R., Jones, S., Jones, T., Jones, U., Jones, V., Jones, W., Jones, X., Jones, Y., Jones, Z., Jones, AA, Jones, AB, Jones, AC, Jones, AD, Jones, AE, Jones, AF, Jones, AG, Jones, AH, Jones, AI, Jones, AJ, Jones, AK, Jones, AL, Jones, AM, Jones, AN, Jones, AO, Jones, AP, Jones, AQ, Jones, AR, Jones, AS, Jones, AT, Jones, AU, Jones, AV, Jones, AW, Jones, AX, Jones, AY, Jones, AZ, Jones, BA, Jones, BB, Jones, BC, Jones, BD, Jones, BE, Jones, BF, Jones, BG, Jones, BH, Jones, BI, Jones, BJ, Jones, BK, Jones, BL, Jones, BM, Jones, BN, Jones, BO, Jones, BP, Jones, BQ, Jones, BR, Jones, BS, Jones, BT, Jones, BU, Jones, BV, Jones, BW, Jones, BX, Jones, BY, Jones, BZ, Jones, CA, Jones, CB, Jones, CC, Jones, CD, Jones, CE, Jones, CF, Jones, CG, Jones, CH, Jones, CI, Jones, CJ, Jones, CK, Jones, CL, Jones, CM, Jones, CN, Jones, CO, Jones, CP, Jones, CQ, Jones, CR, Jones, CS, Jones, CT, Jones, CU, Jones, CV, Jones, CW, Jones, CX, Jones, CY, Jones, CZ, Jones, DA, Jones, DB, Jones, DC, Jones, DD, Jones, DE, Jones, DF, Jones, DG, Jones, DH, Jones, DI, Jones, DJ, Jones, DK, Jones, DL, Jones, DM, Jones, DN, Jones, DO, Jones, DP, Jones, DQ, Jones, DR, Jones, DS, Jones, DT, Jones, DU, Jones, DV, Jones, DW, Jones, DX, Jones, DY, Jones, DZ, Jones, EA, Jones, EB, Jones, EC, Jones, ED, Jones, EE, Jones, EF, Jones, EG, Jones, EH, Jones, EI, Jones, EJ, Jones, EK, Jones, EL, Jones, EM, Jones, EN, Jones, EO, Jones, EP, Jones, EQ, Jones, ER, Jones, ES, Jones, ET, Jones, EU, Jones, EV, Jones, EW, Jones, EX, Jones, EY, Jones, EZ, Jones, FA, Jones, FB, Jones, FC, Jones, FD, Jones, FE, Jones, FF, Jones, FG, Jones, FH, Jones, FI, Jones, FJ, Jones, FK, Jones, FL, Jones, FM, Jones, FN, Jones, FO, Jones, FP, Jones, FQ, Jones, FR, Jones, FS, Jones, FT, Jones, FU, Jones, FV, Jones, FW, Jones, FX, Jones, FY, Jones, FZ, Jones, GA, Jones, GB, Jones, GC, Jones, GD, Jones, GE, Jones, GF, Jones, GG, Jones, GH, Jones, GI, Jones, GJ, Jones, GK, Jones, GL, Jones, GM, Jones, GN, Jones, GO, Jones, GP, Jones, GQ, Jones, GR, Jones, GS, Jones, GT, Jones, GU, Jones, GV, Jones, GW, Jones, GX, Jones, GY, Jones, GZ, Jones, HA, Jones, HB, Jones, HC, Jones, HD, Jones, HE, Jones, HF, Jones, HG, Jones, HH, Jones, HI, Jones, HJ, Jones, HK, Jones, HL, Jones, HM, Jones, HN, Jones, HO, Jones, HP, Jones, HQ, Jones, HR, Jones, HS, Jones, HT, Jones, HU, Jones, HV, Jones, HW, Jones, HX, Jones, HY, Jones, HZ, Jones, IA, Jones, IB, Jones, IC, Jones, ID, Jones, IE, Jones, IF, Jones, IG, Jones, IH, Jones, II, Jones, IJ, Jones, IK, Jones, IL, Jones, IM, Jones, IN, Jones, IO, Jones, IP, Jones, IQ, Jones, IR, Jones, IS, Jones, IT, Jones, IU, Jones, IV, Jones, IW, Jones, IX, Jones, IY, Jones, IZ, Jones, JA, Jones, JB, Jones, JC, Jones, JD, Jones, JE, Jones, JF, Jones, JG, Jones, JH, Jones, JI, Jones, JJ, Jones, JK, Jones, JL, Jones, JM, Jones, JN, Jones, JO, Jones, JP, Jones, JQ, Jones, JR, Jones, JS, Jones, JT, Jones, JU, Jones, JV, Jones, JW, Jones, JX, Jones, JY, Jones, JZ, Jones, KA, Jones, KB, Jones, KC, Jones, KD, Jones, KE, Jones, KF, Jones, KG, Jones, KH, Jones, KI, Jones, KJ, Jones, KL, Jones, KM, Jones, KN, Jones, KO, Jones, KP, Jones, KQ, Jones, KR, Jones, KS, Jones, KT, Jones, KU, Jones, KV, Jones, KW, Jones, KX, Jones, KY, Jones, KZ, Jones, LA, Jones, LB, Jones, LC, Jones, LD, Jones, LE, Jones, LF, Jones, LG, Jones, LH, Jones, LI, Jones, LJ, Jones, LK, Jones, LL, Jones, LM, Jones, LN, Jones, LO, Jones, LP, Jones, LQ, Jones, LR, Jones, LS, Jones, LT, Jones, LU, Jones, LV, Jones, LW, Jones, LX, Jones, LY, Jones, LZ, Jones, MA, Jones, MB, Jones, MC, Jones, MD, Jones, ME, Jones, MF, Jones, MG, Jones, MH, Jones, MI, Jones, MJ, Jones, MK, Jones, ML, Jones, MM, Jones, MN, Jones, MO, Jones, MP, Jones, MQ, Jones, MR, Jones, MS, Jones, MT, Jones, MU, Jones, MV, Jones, MW, Jones, MX, Jones, MY, Jones, MZ, Jones, NA, Jones, NB, Jones, NC, Jones, ND, Jones, NE, Jones, NF, Jones, NG, Jones, NH, Jones, NI, Jones, NJ, Jones, NK, Jones, NL, Jones, NM, Jones, NN, Jones, NO, Jones, NP, Jones, NQ, Jones, NR, Jones, NS, Jones, NT, Jones, NU, Jones, NV, Jones, NW, Jones, NX, Jones, NY, Jones, NZ, Jones, OA, Jones, OB, Jones, OC, Jones, OD, Jones, OE, Jones, OF, Jones, OG, Jones, OH, Jones, OI, Jones, OJ, Jones, OK, Jones, OL, Jones, OM, Jones, ON, Jones, OO, Jones, OP, Jones, OQ, Jones, OR, Jones, OS, Jones, OT, Jones, OU, Jones, OV, Jones, OW, Jones, OX, Jones, OY, Jones, OZ, Jones, PA, Jones, PB, Jones, PC, Jones, PD, Jones, PE, Jones, PF, Jones, PG, Jones, PH, Jones, PI, Jones, PJ, Jones, PK, Jones, PL, Jones, PM, Jones, PN, Jones, PO, Jones, PP, Jones, PQ, Jones, PR, Jones, PS, Jones, PT, Jones, PU, Jones, PV, Jones, PW, Jones, PX, Jones, PY, Jones, PZ, Jones, QA, Jones, QB, Jones, QC, Jones, QD, Jones, QE, Jones, QF, Jones, QG, Jones, QH		

Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moulé, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrett, B.G.

The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*

Nature 400 (6744), 532-538 (1999)

JOURNAL MEDLINE 99376085

PUBMED 10448855

REFERENCE 2 (bases 1 to 86827)

AUTHORS Bowman, S., Skelton, J., Churcher, C., Lawson, D., Quail, M. and Barrett, B.

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 86827)

AUTHORS Lawson, D., Bowman, S. and Barrett, B.

TITLE Direct Submission

JOURNAL Submitted (17-Dec-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT On or before May 14, 2001 this sequence version replaced g1:2982535, g1:2982536, g1:2894454, g1:2982554, g1:2982562, g1:2894489, g1:2982572, g1:2982574, g1:4493931. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum.

FEATURES

source location/Qualifiers

1. 86827

/organism="Plasmodium falciparum 3D7"

/strain="3D7"

/db_xref="taxon:36329"

/chromosome="3"

/clone="MAL3P5"

join(322..603,826..1023,1197..1301,1458..2942)

/gene="PFC0575w, MAL3P5.1"

join(322..603,826..1023,1197..1301,1458..2942)

/gene="PFC0575w, MAL3P5.1"

/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689 aa, possible signal sequence, revised: added new exon 2"

/codon_start=1

/product="Hypothetical protein, PFC0575w"

/protein_id="CAB38969.2"

/db_xref="GI:7711065"

/db_xref="SPTRMBL:O97258"

/translation="MYLKNVYIYISCFILFDLCFPHLLMKYKNNMNNKSTYFELRSPOIYRRKRSRIKRVSPKKOKKPLFENIKKSGSEFGFRNOYDOKITIDVTSNINNTIRIKOYTHKKKSNSTKEMIKQILHCVSKIDFKITNNLSYIKRIFQSDVSHSLNQLISSEKVEKDAENYLAHLHLDENITLPSMHIMDFSKOVCISNIRIKSKRRKNNLSIYNLICTLITFYCMCLLIYISHLCEFFFCFPCVNICIRDIYECVGLIRKIERLYNLCCEKKIKFHKMDAKKPEIMNKDDDLVENVYDELICFTMKLNIRNNNNIIRSNYINNDISIDKDMWNNIDVNNINNSIDEXIKKQREMPDDENIKELDTYFOFOLFNDNIKYIEEDPLVYNINSNNINNNINNTKKNKHTKDYNDDDDIYEEKEEDLVYTKNIDDIYKNTIGNNKSLSEKKNFLEQADIEFNFLSNVNDQGRKVSNDENTSTEHKKNKTKNGYDTLEIQOMENNFIRKINDNISDIISNNEINIIKLKRLNOSDEDINLSDIYERLRTKVLVYLOKLEYLKKFYQYDIINQYPIIKKEKFTVLDLNLGYKIVMSPDVNSLEFKTKIDISIPNEKDKNNQNMOKSKNAVIMENVW"

604..609

/gene="PFC0575w, MAL3P5.1"

/note="predicted splice donor sequence for exon 1 of PFC0575w (revised)"

816..825

/gene="PFC0575w, MAL3P5.1"

/note="predicted splice acceptor sequence for exon 2 (revised of PFC0575w)"

1024..1029

/gene="PFC0575w, MAL3P5.1"

/note="predicted splice donor sequence for exon 2 of PFC0575w (revised)"

complement(3354..6644)

/gene="PFC0580c, MAL3P5.2"

complement(3354..6644)

/gene="PFC0580c, MAL3P5.2"

/note="PFC0580c (MAL3P5.2), Hypothetical protein, len: 1097 aa, possible signal sequence, predicted using hexExon"

/codon_start=1

/product="Hypothetical protein, PFC0580c"

/protein_id="CAB38968.1"

/db_xref="GI:4493932"

/db_xref="SPTRMBL:O97257"

/translation="MGLKKNVFFLLVSPSLFAVFLKRRHNDENNVEYLINNDIEKIKIRIHKKSYIDPLEFNIDYISYIKKILMYLFRKRRKDEEYVYITNMYRKRRRAIKYNFIDSEONLFNFKFYIEVLEISLSPSLISLYLKINSECVNTPSLYYNKLNDKNHNLINNEKLTKEFKOINNEHGOAPNTHHNNNNNNKPIQNTISCRTNYSSTVNNMERKNITDYKYNINHFYDEKSYSDENCOHTLTMVITLNTQDNICICLNTNNTIYITMIMILYICKNNEIILNTNHDMMKKNQNNNTSLEKVIYSEFEKKEENHYDLEDDQMMNLHKKENDKFNYSNENTHNNIYVISONYFDHINSNNSRCSFNKLRKQOTDDTKHLMGKEKYPMKNSDEKKNNTCNINTEKDOKRIILKIYFLKGNKLDIDILNELVYIMYMLFEGSLISIKRNHILKEKMEFDKDKIYILNSADYMMNLRLNRLIKFSKRENERENINSFASPLISKVIYFEEDIGRSRONTIYNNVYDKETNKTNTNNNNNDNDCSNNDHCNNNDHICSNNDNICSNNNNNCSNNNNNCSNNNNNCSNNKMLDERFCDDKPFNDYTRKREKRTVELAKITNNIDYDLGKKKEKHQNDNTNLYIKKFEPMYILNNTINKDTSFIEHNNIIVNGHINKNNNILEKKYNLFESSISYFIKDIYENYKRLVYIDNLIKFCRVELKNNHINRKLVMKRAFHYYIYNFQFIINNYHIIHKNNHRIHHLKCKREIDIVKFKDLYCYMNNNNINFIYIKHVDNECVRFKAYNKLILEYVYLKEKENIYKNNIKKILYITNNNISNDIYVNTSYNKTYNQNKYNNKNSHTEYSLKNDHILFLYTORIQNCCDSESTIYKKNENKPNLNTIYELHITVYSEKKKKKFFSFISSPVSDTVNSFTFSYFSLYFLFLFHPDMYASYIFKTLVYSGLPYYSLVNININVCGRKT"

join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)

/gene="PFC0581w"

join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)

/gene="PFC0581w"

/note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction very tentative"

/codon_start=1

/product="hypothetical protein, PFC0581w"

/protein_id="CAB90285.1"

/db_xref="GI:7711066"

/translation="MKKKKHYSFKPVNDKKNVYLCSDYIRIINMGHITRLQKIHITFANQILASSFESHKNTKVTEDIIDGKWFNTDVLISEIRKQDDFSNDIKTYIVAMKNDQDLTYSNNKIDTVNASFQMAQSPETIFLNKFSHMSGPA LKVDDEELVSKNNESFSLSDNSVYKTYVDTLNDNIEESTKYNFASGVVYTLKREKKTINRLLLSREKYPNMQVMMKREKRIIITFEVTLNLEFELLSHRYHDSVONFLKEERNNDKIQDDIDDEBEKFTDEILREAKKKEEYDKDDEEL"

7790..7799

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 2 of PFC0581w"

7830..7835

/gene="PFC0581w"

/note="potential splice donor sequence for exon 2 of PFC0581w"

7901..7911

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 3 of PFC0581w"

7941..7947

/gene="PFC0581w"

/note="potential splice donor sequence for exon 3 of PFC0581w"

8055..8063

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 4 of PFC0581w"

8107..8112

/gene="PFC0581w"

/note="potential splice donor sequence for exon 4 of PFC0581w"

8293..8301

1097 aa, possible signal sequence, predicted using hexExon"

/codon_start=1

/product="Hypothetical protein, PFC0580c"

/protein_id="CAB38968.1"

/db_xref="GI:4493932"

/db_xref="SPTRMBL:O97257"

/translation="MGLKKNVFFLLVSPSLFAVFLKRRHNDENNVEYLINNDIEKIKIRIHKKSYIDPLEFNIDYISYIKKILMYLFRKRRKDEEYVYITNMYRKRRRAIKYNFIDSEONLFNFKFYIEVLEISLSPSLISLYLKINSECVNTPSLYYNKLNDKNHNLINNEKLTKEFKOINNEHGOAPNTHHNNNNNNKPIQNTISCRTNYSSTVNNMERKNITDYKYNINHFYDEKSYSDENCOHTLTMVITLNTQDNICICLNTNNTIYITMIMILYICKNNEIILNTNHDMMKKNQNNNTSLEKVIYSEFEKKEENHYDLEDDQMMNLHKKENDKFNYSNENTHNNIYVISONYFDHINSNNSRCSFNKLRKQOTDDTKHLMGKEKYPMKNSDEKKNNTCNINTEKDOKRIILKIYFLKGNKLDIDILNELVYIMYMLFEGSLISIKRNHILKEKMEFDKDKIYILNSADYMMNLRLNRLIKFSKRENERENINSFASPLISKVIYFEEDIGRSRONTIYNNVYDKETNKTNTNNNNNDNDCSNNDHCNNNDHICSNNDNICSNNNNNCSNNNNNCSNNNNNCSNNKMLDERFCDDKPFNDYTRKREKRTVELAKITNNIDYDLGKKKEKHQNDNTNLYIKKFEPMYILNNTINKDTSFIEHNNIIVNGHINKNNNILEKKYNLFESSISYFIKDIYENYKRLVYIDNLIKFCRVELKNNHINRKLVMKRAFHYYIYNFQFIINNYHIIHKNNHRIHHLKCKREIDIVKFKDLYCYMNNNNINFIYIKHVDNECVRFKAYNKLILEYVYLKEKENIYKNNIKKILYITNNNISNDIYVNTSYNKTYNQNKYNNKNSHTEYSLKNDHILFLYTORIQNCCDSESTIYKKNENKPNLNTIYELHITVYSEKKKKKFFSFISSPVSDTVNSFTFSYFSLYFLFLFHPDMYASYIFKTLVYSGLPYYSLVNININVCGRKT"

join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)

/gene="PFC0581w"

join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)

/gene="PFC0581w"

/note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction very tentative"

/codon_start=1

/product="hypothetical protein, PFC0581w"

/protein_id="CAB90285.1"

/db_xref="GI:7711066"

/translation="MKKKKHYSFKPVNDKKNVYLCSDYIRIINMGHITRLQKIHITFANQILASSFESHKNTKVTEDIIDGKWFNTDVLISEIRKQDDFSNDIKTYIVAMKNDQDLTYSNNKIDTVNASFQMAQSPETIFLNKFSHMSGPA LKVDDEELVSKNNESFSLSDNSVYKTYVDTLNDNIEESTKYNFASGVVYTLKREKKTINRLLLSREKYPNMQVMMKREKRIIITFEVTLNLEFELLSHRYHDSVONFLKEERNNDKIQDDIDDEBEKFTDEILREAKKKEEYDKDDEEL"

7790..7799

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 2 of PFC0581w"

7830..7835

/gene="PFC0581w"

/note="potential splice donor sequence for exon 2 of PFC0581w"

7901..7911

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 3 of PFC0581w"

7941..7947

/gene="PFC0581w"

/note="potential splice donor sequence for exon 3 of PFC0581w"

8055..8063

/gene="PFC0581w"

/note="potential splice acceptor sequence for exon 4 of PFC0581w"

8107..8112

/gene="PFC0581w"

/note="potential splice donor sequence for exon 4 of PFC0581w"

8293..8301

[illegible]

QY- 455 TTGATTACTATGAGATTCTTCTTACCAATTTAAATTCGAATTCATATCTGTATAT 514
Db 39259 AATTAATAATATATATATATTTATTAATAAATAAAT --- AATTAATAATATATATATAT 39315
QY 515 TAAATTAATACGAGTGTGAATAT -CCGTTATTCGATTCACCTCCAAATCATGATTAATGATTC 573
Db 39316 TAAATTAATAATTAATTAATATATATATATATTTATTAATAAATAAATAAATTAATATATAT 39375
QY 574 TTGCTGAATCCAGCAAAATTTATTAACAAGTGTGAGAAAACCGAAAATAGAAAAG 633
Db 39376 ATTATTAATAATAATAATAATTAATATATATATATTTATTAATAAATAAATAAATAATA 39435
QY 634 GGAAGGTGTGACCAATGAGACTAGTCAATTAATTAATCAAGAAATTAAGAGATGACAA 693
Db 39436 TATATTTATTAATAATAATAATAATAATAATTAATTAATTAATTAATAATAATAATAATA 39495
QY 694 CCAAAAGTTGTGGAATTAATGTCCCTGACACTTTCCTCAGCATTAATCGACCTTA 753
Db 39496 TTATATTAATAATAATAATAATAATTAATTAATTAATTAATTAATAATAATAATAATAAT 39555
QY 754 TTGGATTTCTCGAATATTCGTTAAATTTGCGATACGATGTGAAAAATTTTATAT 813
Db 39556 AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 39611
QY 814 GTTAGCTGATTCGAATTAATGTGTCCAGTATTTGCAATATCTCTGTTTAAAGCATAT 873
Db 39612 AATTAATTCGAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 39671
QY 874 TTGCTCTCTTTTCTTCGTTCTCTTAACATATATTAATTCGGGATATATATGATACAA 933
Db 39672 AATCTTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 39731
QY 934 TGAATATACAAAAACAATGTGTGGGCAATTTTGAATAACTTTTTCGAAACATTAAC 993
Db 39732 ATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 39791
QY 994 GGGACACGAGCTGACCCCTTAAATACGATTTTACAGCTGACGTGAGATTTCTAG 1053
Db 39792 TAAATTAATTTTATTTTATTTTATTCACCTTAATAATATTAATAATAATAATAATAATAAT 39851
QY 1054 CATTAACATTAAGACCCGTTCAAGCTTTTATACAAAGTTTCAACAGTAATATAGCTT 1113
Db 39852 AATATTTTAAATAATATATTAATTAATAATTTATTTTATTAATAATAATAATAATAACAT 39911
QY 1114 GAATCCTTAGAAAATTTTGCAATTAACGGTGTATATTAATAATTAAGATTAGTGGTAA 1173
Db 39912 TTTATATCATTAATTAATTTCTTAATTAATAATAATAATAATAATAATAATAATAATAATAAT 39968
QY 1174 ACAATATGTTAATCATTAAGTGTGACAAATATACATAATTCCTTACAGAAAACAAC 1233
Db 39969 ATTAATAATAATGTTTCTTATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 40028
QY 1234 TTAAAGAGTATACATATCCATTAATGGGATGCTATACCTTACAGTATGATTAATA 1293
Db 40029 AATATTAATTAATTAATTAATTAATTAAT -TTATTAATAATAATAATAATAATAATAATAATAATAATAATA 40087
QY 1294 GAGACTAAGAACTTATGTATGTGATGCAATAATAATTCACACCGGTGTAATAATTA 1353
Db 40088 ATTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1413
QY 1354 TGGGACCGTATGTCAGTACCTGCAAAATATCATCTTGGTGTGTCACAAATTAATAACAA 1413
Db 40148 TTTTAAATTAATTAATGAAATCAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 40207
QY 1414 AAACAAGAAAAAAGAAAAAGATTTTCTTGGATTCATTCATGATCTRAAAATGATAG 1473
Db 40208 GTATTAATTAATTAATGTTTAATTTATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATA 40267
QY 1474 ATCTTTGGGTAAAGTTTGAAGTCCCTTCAAGAGGTGAACAGTCACTGACATTAATAA 1533
Db 40268 AATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 40327
QY 1534 TTGCTTCTTAAATGATCTTTAAACATTTATTTATGTAATTTAATAAGACGAA 1593

Db	40328	TATATATTAATTAATTAATTTTA-TTATTTTAAATTAATTAATTAATTTTAATTA	40386
Oy	1594	CTGTACATTTACCAATTTTATATGTACGTACGTATGATTAAT--TCCAATACATAC	1651
Db	40387	TTTTTTTACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40446
Oy	1652	TTTGGATGTTAAACCTTAATCTGTGTTCTCCACGCTGAATTAATTAATTAATTAATTA	1711
Db	40447	TTTTTAATTAATTAATTTTAATTTAATTTATTTATTTACTTTTAAGTATTTTAATTTTA	40506
Oy	1712	AAAAAGTTTTGTCCTAATTTTCGCATGATGAAGATAACCTAATGACTTTAATTTT	1771
.Db	40507	ATATTTATTAATTAATTAATTTATTAATTAATTAATTAATTAATTTAATTAATTTA	40566
Oy	1772	TGAAATGTAACCTTTTACT---CATGATTAATTAACCGTATTTTGTGGCATA	1827
Db	40567	ATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40626
Oy	1828	TGACAGCCTGACACGTGA-----TAGTCATTTTTCGCAAAATTAATTAATTAAG	1880
Db	40627	TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTA	40686
Oy	1881	AATTCATGCTACTATCAATGAGAAACAGCTGAGTATTAATTAATTTAATTAAGACA	1940
Db	40687	ATTAATAATTAATTTTAATTTAATTAATTAATTAATTTAATTTAATTTAATTTAATA	40746
Oy	1941	AATTTTGGAAAAATGTTATTAATTTGACAAATTAATTAATAATGATGCTTAATGTA	2000
Db	40747	ATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA	40806
Oy	2001	TTTCCATGTCCTAAAAATTTTCTTTTATTAATTTAGTATTAATTAATTAATTAAGACAT	2060
Db	40807	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40864
Oy	2061	AATAGTTGCTGAATTCCAATATCTCCATTAATTTTGTGAATCTACAAATTAATTAATA	2120
Db	40865	TATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40924
Oy	2121	TTTATGCAATTAACAAATGATAGAAAGTCCAAAAAAATTTGTTAACAGAAATCTCCAA	2180
Db	40925	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40984
Oy	2181	ATTTTCTTTTAAATGGAACAAGAAATTAACAGATAGAAACATATTTGTTGGAAATGGA	2240
Db	40985	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	41044
Oy	2241	AGTAGAATTAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2300
Db	41045	TTTAAATTAATTTAATGAATTAATTTAATTAATTTAATTAATTTAATTAATTTAATTT	41104
Oy	2301	ATGTTATCTGATGAGGTAAATTAATTAATGATGATGCGATTCAGAAATGGGACAACAATG	2360
Db	41105	ATATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	41163
Oy	2361	AAAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2407
Db	41164	TATATTTAATTAATTTATTAATTAATTTAATTAATTTAATTAATTTAATTAATTAATTA	41210

RESULT 11					
AC120883/c	AC120883	139877 bp	DNA	linear	HTG_06-AUG-2002
LOCUS					
DEFINITION	Homo sapiens chromosome 18 clone CTD-2146H24 map 18, WORKING DRAFT				
ACCESSION	AC120883				
VERSION	AC120883.2	GI:22123189			
KEYWORDS	HTG; HTGS_PPHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 139877)				

AUTHORS
 TITLE
 JOURNAL
 REFERENCE

Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone CTD-2146H24
 Unpublished
 2 (bases 1 to 133877)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Camerata, J., Campilano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Grude, S., Gordon, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Haxton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., Larocque, K., Lamares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., Mckernan, K., Meldrum, J., Meneses, L.,
 Milnova, T., Milenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Notman, C.H., O'Connor, T., O'Donnell, P., O'Mell, D.,
 Oliver, J., Peterson, C., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Rothmann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
 Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-May-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 133877)
 Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,

TITLE
 JOURNAL
 COMMENT

Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
 Camarata, J., Chang, J., Chazaro, R., Choepel, Y., Collamore, A.,
 Cook, A., Cooke, P., DeArllano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Foo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galaan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, R., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meidrim, J., Menus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicoli, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Plunkhang, P., Plerre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schnpbach, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (26-Aug-2002) Whitehead Institute/MIT Center for Genome
 Research, 330 Charles Street, Cambridge, MA 02141, USA
 On Aug 6, 2002 this sequence version replaced gi:20531967.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26769
Center clone name: 2146_H-24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132654 bases at least Q40
Consensus quality: 132863 bases at least Q30
Consensus quality: 133065 bases at least Q20
Insert size: 133000; agarose-fp
Insert size: 133577; sum-of-ctrls
Quality coverage: 17.1 in Q20 bases; agarose-fp

```



```
/note="assembly_name:Contlg111"
```

b

	843	111GCAAACTTCGGTTTAAAGCATATTTGTCTTCCTTTGTTTGCTTCCTTA	902
	78994	ATCATATTATACATAATTACATAACTATATATCATAATATTATATATTACGTAACATAAT	78935


```

repeat_unit      1706..2043
                  /note="repeat I-B2"
                  /rpl_type=tandem
repeat_unit      2044..2388
                  /note="repeat I-C"
                  /rpl_type=tandem
misc_feature      2491..2511
                  /note="deoxythymidylate stretch"
repeat_unit      2512..2648
                  /partial
repeat_unit      2649..3112
                  /note="Repeat II-A"
                  /rpl_type=tandem
repeat_unit      3113..3576
                  /note="Repeat II-B1"
                  /rpl_type=tandem
repeat_unit      3577..4040
                  /note="Repeat II-B2"
                  /rpl_type=tandem
repeat_unit      4041..4304
                  /note="Repeat II-C"
                  /rpl_type=tandem
misc_feature      complement(4565..4585)
                  /note="deoxythymidylate stretch"
BASE COUNT      2271 a 131 c 74 g 2125 t
ORIGIN

```

```

Query Match      3.98; Score 124.4; DB 3; Length 4601;
Best Local Similarity 43.98; Pired. No. 6.3e-09;
Matches 1027; Conservative 0; Mismatches 1276; Indels 37; Gaps 10;

66 TGTTCAACATGATTGGTTGGTTCATTTGGATTCAGATTAATTAATTAAT 125
2278 TATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 125
126 TCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2337
2338 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 185
186 CTAATCTTGGTATATAGAGAAATATGAGTTGGATTTATTAATTAAT 2397
2398 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2455
246 GATTCATTTGGTGGATTAACAGTTAGTTTGTCTTCTTGTATATGAT 305
2456 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2515
306 AGTAAATCAAAAAGATTAATGATGAGTGAACATATTCGTTATGACCC 365
2516 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2575
366 AAAAAAACAACAACAACAACCCCGATATAGTTTGTGTTGGATTAAGT 425
2576 ATATATATACAAAATTA-----AATTAATTAATTAATTAATTAAT 2620
426 TATTTGATCAATTAATTAATGATGATCAATTTCTTGAATTAATTAAT 485
2621 TTTTAAAAATTTCTTAATGATTAATTTTATTAATTAATTAATTAAT 2680
486 AAAATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 545
2681 TTTTAAAAATTAACAAAAATTTTATTAATTAATTAATTAATTAATTAAT 545
546 TCGATCAGCAATCATGATTAATGATTCCTGTGCTAATCCGAAATTAAT 2740
2741 TATTTTCAATTTTAAAAAATTTTAAAAAATTAATTTTAAAAAATTAAT 605
606 ATTTGAGAAAAACCAAAATTAAGAAAGGAAAGATGACCATGAGATATG 2800
2801 CTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 665
666 AATTAATCAAGATTAAGATGACCAACAAAAGTTGTGATATATATGCTCC 2860

```

```

2861 ATATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2920
724 AGCTTCTCTCAACATCAATTAATGACCCATTTGATTTGCGATTAATTAAT 783
2921 AATTTATATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2980
784 GCGATACAGATTTGTGAAA--AATTTATTTATTTGTTAGCTGATCAATTAAT 840
2981 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3040
841 GGTATTTGCAATATCTCTGTTTAAAGCATTTTGTCTTTCTTTGCTGCTCT 900
3041 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 955
901 TAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1015
3101 AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1075
956 CTGGGACCATTTTGAATTAATCTTTTCTCAACATTAAGGACATGACCTGTA 1130
3161 AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1190
1016 AATACGATTTTACAGGCTACATTAATTAATTAATTAATTAATTAATTAATTAAT 1250
3221 TTTAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 1310
1076 CA-----AGTATTTATTAACAAAGTTTAAAGTAAATTAATTAATTAATTAAT 1367
3281 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1427
1131 TTTGAAATTAACGGTGTATATTAATTAATTAATTAATTAATTAATTAATTAAT 1487
1191 TTAGTGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1546
3401 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1606
1251 ATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1666
3461 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1726
1311 ATGATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1786
3521 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1846
1368 AGATTCACCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1906
3581 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1966
1428 GAAACGATTTTCTTGGATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 2026
3641 ATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2086
1488 AGTTTGAAGCTCTACAGGCTGTAACCTGCAATTAATTAATTAATTAATTAATTAAT 2146
3701 AAAAAATTAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAA 2206
1547 TGCATCTTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2266
3761 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2326
1607 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2386
3821 AATTTAAAAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2446
1667 TTAATCTGTTTCTTCAAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2506
3881 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2566
1727 TATTTTGGCGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2626

```

Db 3941 ATTATGATATTTATATATATATATAT -ATATAGAAAAATTAATTTATTAATTA 3999

Qy 1787 TTTACTATGATTAATTAATACCGTATGTTTGGCCATATAGACGCTCACAACGT 1846

Db 4000 TTTAT 4059

Qy 1847 GATAGTCATATTTTTCGCAATATTAATTAATGATGATGATGATGATGATGATGATG 1906

Db 4060 AATTAAT 4119

Qy 1907 AACAGCTGATTAAT 1966

Db 4120 AAT 4179

Qy 1967 AACAT 2026

Db 4180 TTTTAAAAAAACCTAT 4239

Qy 2027 TTTTAT 2086

Db 4240 TTTAAT 4299

Qy 2087 ATTAAT 2146

Db 4300 AAT 4359

Qy 2147 TTTCAAAAAAT 2206

Db 4360 TAAAT 4419

Qy 2207 TTAACAT 2266

Db 4420 TAT 4479

Qy 2267 TAAAT 2326

Db 4480 AAAAAAT 4539

Qy 2327 ATGCATGTCGATTCAGATATGAGCAATGAAACGAAATTAATTAATCTTA 2386

Db 4540 TTTTAAAAAAATAGTAAATATATATATATATATATATATATATATATATATATAT 4599

RESULT 15

LOCUS DMU37541 19517 bp DNA circular INV 02-MAR-2001

DEFINITION Drosophila melanogaster complete mitochondrial genome.

ACCESSION U37541

VERSION U37541.1 GI:1166529

KEYWORDS

SOURCE Drosophila melanogaster.

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 12511 to 12682) Clary,D.O., Goddard,J.M., Martin,S.C., Faaron,C.M. and Wolstenholme,D.R.

AUTHORS

TITLE Drosophila mitochondrial DNA: a novel gene order

JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)

MEDLINE 83090428

PUBMED 6294611

REFERENCE 2 (bases 5269 to 5695) Clary,D.O., Wahlthner,J.A. and Wolstenholme,D.R.

AUTHORS

TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes

JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)

MEDLINE 83220794

PUBMED 6304652

REFERENCE 3 (bases 404 to 5272) de Bruijn,M.H.

AUTHORS

TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and

JOURNAL genetic code

MEDLINE Nature 304 (5923), 234-241 (1983)

PUBMED 83245048

REFERENCE 4 (bases 804 to 1778) Satta,Y., Ishiwa,H. and Chigusa,S.I.

AUTHORS

TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species

JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)

MEDLINE 88174373

PUBMED 2832697

REFERENCE 5 (bases 5268 to 13619) Garesse,R.

AUTHORS

TITLE Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations

JOURNAL Genetics 118 (4), 649-663 (1988)

MEDLINE 88212147

PUBMED 3130291

REFERENCE 6 (bases 441 to 2967) Satta,Y. and Takahata,N.

AUTHORS

TITLE Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)

MEDLINE 91088557

PUBMED 2124697

REFERENCE 7 (bases 14215 to 14512) Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.

AUTHORS

TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods

JOURNAL Science 258 (5086), 1345-1348 (1992)

MEDLINE 93088057

PUBMED 1455227

REFERENCE 8 (bases 14917 to 19517) Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.

AUTHORS

TITLE Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA

JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)

MEDLINE 94285822

PUBMED 8015445

REFERENCE 9 (bases 1 to 408; 13319 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.

AUTHORS

TITLE Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons

JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)

MEDLINE 96423163

PUBMED 8825764

REFERENCE 10 (bases 1 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.

AUTHORS

TITLE Direct Submission

JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, State University, East Lansing, MI 48824-1319, USA

FEATURES

SOURCE location/Qualifiers

1..19517

/organism="Drosophila melanogaster"

/organelle="mitochondrion"

/db_xref="taxon:722"

/note="derived from new and previously submitted sequences: sequence is a composite containing sequences obtained from different Drosophila melanogaster strains"

1..65

/product="tRNA-Ile"

/complement(97..165)

/product="tRNA-Gln"

171..239

/product="tRNA-Phe"

240..1265

/codon_start=1

/transl_table=5

/product="NADH dehydrogenase subunit 2"

/protein_id="AAC47811.1"

/db_xref="GI:1166530"

/translation="MFNNSKILFTITMIGTILITVSNLWGLAMGLEINLSEFIP

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 10, 2003, 21:59:33 ; Search time 733.005 Seconds

(Without alignments)
9837.455 Million cell updates/sec

Title: US-09-502-426B-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atgtggtattatattgtgtg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3202	100.0	6888	21	AAA59599
2	84.6	2.6	883	22	AA15210
3	84.6	2.6	9810	24	ABK32427
4	83.8	2.6	8170	24	ABK28257
5	82.8	2.6	6131	24	ABL32890
6	82.6	2.6	9810	24	ABL32426
7	82.6	2.6	15732	22	AA54388
8	82.6	2.6	15732	24	ABK28233
9	81.6	2.5	5689	22	AA545384

C	10	81.6	2.5	5689	22	AA546426	Tumour suppressor
C	11	81.6	2.5	5689	24	ABK28226	DNA transcription
C	12	80.6	2.5	6294	24	ABL33054	Human immune syste
C	13	80.4	2.5	6131	24	ABL32891	Human immune syste
C	14	80	2.5	64081	24	ABK32787	Buchnera sp. genom
C	15	79.6	2.5	47108	24	ABK31511	Signal transductio
C	16	79.4	2.5	883	22	AA15210	Human breast cance
C	17	79.4	2.5	8310	20	AA229911	CDNA encoding a SC
C	18	79.2	2.5	18283	24	ABL70502	Chemically treated
C	19	79.2	2.5	18283	24	AA561363	Human gene regulat
C	20	78.2	2.4	50000	24	ABL5643	AMEPV genome fragm
C	21	77.8	2.4	17137	24	ABL32191	Human immune syste
C	22	77.6	2.4	9810	24	ABL32427	Human immune syste
C	23	77.4	2.4	19124	18	AA72882	Plasmodium var-7 p
C	24	77.4	2.4	19124	21	AA298287	Plasmodium var-7 p
C	25	77.2	2.4	15373	24	ABL32467	Human immune syste
C	26	77	2.4	11691	24	ABL34241	Human immune syste
C	27	76.8	2.4	50000	24	ABL5643	AMEPV genome fragm
C	28	76.4	2.4	9810	24	ABL32426	Human immune syste
C	29	76.2	2.4	10254	24	ABL33075	Human immune syste
C	30	76.2	2.4	15548	24	ABL34155	Human immune syste
C	31	76	2.4	50000	24	ABL56202	AMEPV genome fragm
C	32	75.6	2.4	6050	24	ABL34011	Human immune syste
C	33	75.2	2.3	5979	22	AA545313	Human immune syste
C	34	75.2	2.3	5979	24	ABK28152	Chemically pretrea
C	35	75	2.3	38342	22	AA546746	DNA transcription
C	36	75	2.3	38342	24	ABK31507	Tumour suppressor
C	37	74.8	2.3	6963	24	ABL32979	Signal transductio
C	38	74.8	2.3	10710	24	ABL32893	Human immune syste
C	39	74.8	2.3	11422	24	ABK39936	Human immune syste
C	40	74.8	2.3	11422	24	ABL32118	Human immune syste
C	41	74.4	2.3	10250	24	ABN80083	Human chemically m
C	42	74.2	2.3	700	22	AAH93026	Human inflammatory
C	43	74.2	2.3	6641	24	ABL54336	Chemically treated
C	44	74.2	2.3	6641	24	ABL33315	Human immune syste
C	45	74	2.3	2104	13	AAQ25273	Sequence encoding

ALIGNMENTS

RESULT 1	AAA59599	standard: DNA; 6888 BP.
ID	AAA59599	
XX	14-NOV-2000	(first entry)
AC	AAA59599;	
XX		
DT		
DE	DNA encoding a cytochrome P450 enzyme designated DMF4.	
KW	DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	plant phenotype; cell elongation; ss.	
XX		
OS	Arabidopsis sp.	
EH	Key	Location/Qualifiers
FT	promoter	1..3203
FT	TATA_signal	3060..3125
FT		/*tag= b
FT	CDS	3203..6110
FT		/*tag= c
FT		/*product= "DMF4"
FT		/*note= "contains introns"
FT	exon	3203..3423
FT		/*tag= d
FT	intron	3424..3503
FT		/*tag= e
FT	exon	3504..3828
FT		/*tag= f
FT	intron	3829..3913
FT		/*tag= g

FT exon 3914..4066
 FT /tag- h
 FT Intron 4067..4164
 FT /tag- i
 FT exon 4165..4479
 FT /tag- j
 FT Intron 4480..4631
 FT /tag- k
 FT exon 4632..4724
 FT /tag- l
 FT Intron 4725..4815
 FT /tag- m
 FT exon 4816..4894
 FT /tag- n
 FT Intron 4895..5000
 FT /tag- o
 FT exon 5001..5110
 FT /tag- p
 FT Intron 5111..5864
 FT /tag- q
 FT exon 5865..6110
 FT /tag- r
 FT 3' UTR 6011..6468
 FT /**tag- s
 PN WO200047715-A2.
 PD 17-AUG-2000.
 PE 11-FEB-2000; 2000WO-US03820.
 PR 11-FEB-1999; 99US-0119657.
 PR 11-FEB-1999; 99US-0119658.
 PA (ARIZ-) ARIZONA BOARD OF REGENTS.
 PI Aspiroz R, Choe S, Feldmann KA;
 DR WPI: 2000-549142/50.
 DR P-PSDB: AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 PT plants, for diagnostic assays and in the production of antibodies -
 PS Claim 3; Fig 10A-G; 113pp; English.
 XX
 CC The present sequence encodes a DMF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
 CC polynucleotide is used for altering the phenotype of a plant. DMF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 Query Match 100.0%; Score 3202; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 AAAATTCATTAAAAATTTTACAGTAATTAATTAATCTTACATGTATTTGTAATACAA 180
 Qy 181 AATATCTATCTTGGTATATGAGAAATATGAGTTTGGAAATTTAATTAATTAAGGANA 240
 Db 181 AATATCTATCTTGGTATATGAGAAATATGAGTTTGGAAATTTAATTAATTAAGGANA 240
 Qy 241 TAATGATTCATTTGGTTGGATTAACAGATTAACTTTTGTCTTTTGTATATGT 300
 Db 241 TAATGATTCATTTGGTTGGATTAACAGATTAACTTTTGTCTTTTGTATATGT 300
 Qy 301 ATATGAGTAATTCAAAAAGAGATTAAGTTGAAGTAAACATATTCGTTAGACCCCA 360
 Db 301 ATATGAGTAATTCAAAAAGAGATTAAGTTGAAGTAAACATATTCGTTAGACCCCA 360
 Qy 361 AAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTC 420
 Db 361 AAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTCAAAAAATTC 420
 Qy 421 AGTTTATTTGATCATTAATTAACATCATTTCTTGTATTAATTAATTAATTAATTC 480
 Db 421 AGTTTATTTGATCATTAATTAACATCATTTCTTGTATTAATTAATTAATTAATTC 480
 Qy 481 CAATTAATTAATTCGAATTCATATCTCTGATTTATTAATTAATTAATTAATTAATTC 540
 Db 481 CAATTAATTAATTCGAATTCATATCTCTGATTTATTAATTAATTAATTAATTAATTC 540
 Qy 541 GTTATGATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 600
 Db 541 GTTATGATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 600
 Qy 601 AGAGTATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAG 660
 Db 601 AGAGTATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAG 660
 Qy 661 TGAATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAA 720
 Db 661 TGAATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAA 720
 Qy 721 GCCAGCTTCTCTCAATCAATATGAGACCCATTTGATTTGATTTGATTTGATTTG 780
 Db 721 GCCAGCTTCTCTCAATCAATATGAGACCCATTTGATTTGATTTGATTTGATTTG 780
 Qy 781 TTTGGGATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAA 840
 Db 781 TTTGGGATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAA 840
 Qy 841 GGTATTTTCATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 900
 Db 841 GGTATTTTCATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 900
 Qy 901 TAACATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 960
 Db 901 TAACATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 960
 Qy 961 ACCATTTTGAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1020
 Db 961 ACCATTTTGAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1020
 Qy 1021 CGATTTTACAGCTCAGTATGATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1080
 Db 1021 CGATTTTACAGCTCAGTATGATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1080
 Qy 1081 TATTTATTAACAAAGTTTACAACTGAATATAGCTTGAATATTCCTTTGAAATTT 1140
 Db 1081 TATTTATTAACAAAGTTTACAACTGAATATAGCTTGAATATTCCTTTGAAATTT 1140
 Qy 1141 CCGGTTGTTATGTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAG 1200
 Db 1141 CCGGTTGTTATGTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAG 1200
 Qy 1201 ACATATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1260
 Db 1201 ACATATTAATTAATTCAGAAAAAATTCAGAAAAAATTCAGAAAAAATTCAGAAAA 1260

QY 1937 ACAAATTTTGGAAAAAGCTTATATCTTCAACATATATTAATAATGATGCGCTATTA 1996
 Db 488 TATTTTGTATTAATAATAGTGTATATTTTGTGTAATTTTGTGAAAT 547
 QY 1997 TGTATTTCCATGTTCTTAAATATTTTATTTTATTTTATTTTATTTTATTTTAA 607
 Db 548 TTTTATTTAAATAGTATTTGTTTATTTAGTATTAATATTTATTTTATTTTATTTTAA 607
 QY 2057 CAATATAGTGTGGAATTCATATTCCTATTAATTTTGGAAATCTACAAATATTT 2116
 Db 608 TATGATTTTATTTTATTAATATAATTTAT-ATTTTATTTTATTTTAAATTTTATGTTA 665
 QY 2117 AATATTTAGCAATACAAATCCATAGAAAGTCCAAATAATTTGTTACGAAACTT 2176
 Db 666 AATATATATCTATGATTTGTTGAGATGCTTAAATATGAAATATG-AAAATTAATTTG 723
 QY 2177 CCAATTTTATTTTATTTTATGACACAAATTAACAGATAGCAAAATTTTGTGGA 2236
 Db 724 GATTTATTTATGTTTATTTTATTTTATTTATTTATTTATTTATTTGTTTATTTT 783
 QY 2237 TCGAGATGATTAATACATTTACCAATTTTAAATAATATATACCTATACGCGCTCA 2296
 Db 784 ATGATATTTAAATATTAATATTAATTTATGCT-ATTTAGTATTAATATGTTT 838
 QY 2297 AAGTATGATATCAGTAGGCTATTAATTAATGATGCTGATTCAGATTTGGGCAAC 2356
 Db 839 AATTTTATTTATGTTTGAATTTTATTTATTTATTTATTTTATTTTGAATTTGAT 898
 QY 2357 AATGAAACGGAATTAATTAATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTT 2416
 Db 899 GAAAAAAGGGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 958
 QY 2417 TCGACTATTTGAGGCAAAAAAAGACAAATGCCAAA 2454
 Db 959 GATTAAGAAAAAATAGTAAATTAATTTATTTGAAAGA 996

RESULT 6
 ABL32426
 ID ABL32426 standard; DNA; 9810 BP.
 AC ABL32426;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 399.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-HIV; anticonvulsant; ophthalmological;
 KW antileukemic; anti-HIV; anticonvulsant; ophthalmological;
 KW antileukemic; anti-HIV; anticonvulsant; ophthalmological;
 KW acute myeloid leukemia; cancer; eye disease; arteriosclerosis; anaemia;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1033529.
 XX
 PA (EPIC-) EPIDEMIOLOGY AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K.
 XX
 DR WPI; 2002-130909/17.

XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1: SEQ ID NO 399; 32pp + Sequence Listing; German.
 CC
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 SQ
 Sequence 9810 BP; 2770 A; 79 C; 1978 G; 4983 T; 0 other;
 Query Match 2.6%; Score 82.6; DB 24; Length 9810;
 Best Local Similarity 46.9%; Pred. No. 0.00031;
 Matches 374; Conservative 0; Mismatches 404; Indels 19; Gaps 3;

QY 1625 TAGATGATGATTTTCCAAATACATTTACTGATGATTTAACTAATCTTTCTTCT 1684
 Db 822 TTTGTTGTAATGTTTGTGTAATGATTTTCTTTCTTTTAAACGTTATTTTATTTT 881
 QY 1685 ACGTATTAATTAATTAATCAATGAGTAATAA--AAGTTTGTCTATTTTGGCATGCA 1741
 Db 882 GTGAATTTTATTTATTTTGTGTTTAAATGTAATTTTATTTTATTTTATTTTATTT 941
 QY 1742 TGAAGATTAACCTTAATGATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTT 941
 Db 942 TTTTATTTTATTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1001
 QY 1802 ATTAACGTAATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1861
 Db 1002 AATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1061
 QY 1862 CTGCAAAATTAATTAATGATTTCAATGCTACTACTACTACTACTACTACTACTACTACT 1921
 Db 1062 TAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1121
 QY 1922 CATTTTAATTTTAAACCAAAATTTTGAATA-----ATGATTAATTTCT 1966
 Db 1122 TTTTATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1181
 QY 1967 AACATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2026
 Db 1182 ATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1241
 QY 2027 TTTTATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2086
 Db 1242 ATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2101
 QY 2087 ATTAATTAATTTTGAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2146
 Db 1302 ATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2161
 QY 2147 TTTCAAAAAAATTTTGTGTTACGAAATTTTCAAAATTTTATTTTATTTATTTATTTAT 2205
 Db 1362 ATTAATTAATTTTGAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1421
 QY 2206 ATTAACAGATTAAGAAATTTTGTGTTGGAATGGAATTTTATTTATTTATTTATTTAT 2265
 Db 1422 AATTAATTAATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1481
 QY 2266 TTTAAAAAATTTTAAAGCTTAATGCTCAAAAGTATTTATTTATTTATTTATTTATTTAT 2325
 Db 1482 ATGATTAATTAATTAATTTGTTTAAAGATTAATTTATTTATTTATTTATTTATTTATTT 1541
 QY 2326 AATGATGTTGATTTCAAAATTTGGGACAAATTAAGCAATTAATTAATTTATTTATTT 2385
 Db 1542 GTAAATTAATTAATTAATTTGTTTAAATTAATTAATTTATTTATTTATTTATTTATTT 1601

ID	ABK28233	standard; DNA; 15732 BP.
XX		
AC	ABK28233;	
XX		
DT	23-APR-2002	(first entry)
XX		
DE	DNA transcription associated genomic DNA #54.	
XX		
KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer; PNA; cytosine methylation state; SNP; retroviral infection; gene; ds; single nucleotide polymorphism; adenosine deaminase deficiency; cancer; viral infection; Sezary syndrome; haematological disorder; tuberculosis; immunological disorder; Werner syndrome; developmental disorder; psoriasis; Rieger's syndrome; neurological disorder; neurodegenerative disorder; Waardenburg syndrome; myelodysplastic syndrome; myocardial infarction; hypertension; arthritis; myelodysplastic disorder; congenital heart disease; polyautamine disorder; solid tumour.	
OS	Unidentified.	
XX		
PN	WO200192565-A2.	
XX		
PD	06-DEC-2001.	
XX		
PF	06-APR-2001; 2001WO-EP03973.	
XX		
PR	06-APR-2000; 2000DE-1019058.	
XX		
PR	07-APR-2000; 2000DE-1019173.	
XX		
PR	30-JUN-2000; 2000DE-1032529.	
XX		
XX	01-SEP-2000; 2000DE-1043826.	
PA	(EPIC-) EPIGENOMICS AG.	
XX		
PI	Olak A, Piepenbrock C, Berlin K;	
XX		
DR	WPI; 2002-090046/12.	
XX		
PT	New nucleic acids or oligomers, useful for diagnosing or treating	
PT	diseases associated with DNA transcription, e.g. immunological	
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid	
PT	tumours or cancer	
XX		
XX	Claim 1: SEQ ID NO 107; 32pp; English.	
XX		
CC	The invention relates to a nucleic acid, which comprises a segment of the	
CC	chemically pretreated DNA of genes associated with DNA transcription from	
CC	one of 346 sequences, and an oligomer, in particular an oligonucleotide	
CC	or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical	
CC	to the chemically pretreated DNA of genes associated with DNA	
CC	transcription. The set of oligomer probes are useful for detecting the	
CC	cytosine methylation state and/or single nucleotide polymorphisms (SNPs)	
CC	in a chemically pretreated genomic DNA. The nucleic acids are useful for	
CC	diagnosing or treating diseases associated with DNA transcription	
CC	(particularly with the methylation status), e.g. adenosine deaminase	
CC	deficiency, viral infection, retroviral infection, Sezary syndrome,	
CC	haematological disorders, immunological disorders, Werner syndrome,	
CC	tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,	
CC	neurological disorders, neurodegenerative disorders, Waardenburg	
CC	syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial	
CC	infarction, hypertension, angiodysplastic syndrome, congenital heart	
CC	disease, HDR syndrome, arthritis, polyautamine disorders, solid tumours	
CC	or cancer. Sequences ABK28127-ABK28472 represent DNA transcription	
CC	associated genomic DNA molecules of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification but was obtained in electronic format directly from the	
CC	European Patent Office.	
XX		
XX		
SO	Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 other;	
XX		
Query Match	2.6%;	Score 82.6; DB 24: Length 15732;
Best Local Similarity	43.9%;	Pred. No. 0.00032;
Matches 433; Conservative	0;	Mismatches 574; Indels 6; Gaps 2;

QY	1400	AACATATAAAAACAAAAACAGAAAAAAGAAAAAGCATTTTCTTGATTCGATTCAATGA	1
Db	6832	AAATATATAAAAACAAAAAATAAAAAAATCAACAAATACTTTATATATTTTTCAT	6
QY	1460	TCTAAATATGATGATCTTTGGCTTACAGTTGGAAGTCCTGACAAAGCGGTAAACAT	6
Db	6772	TCTTAAATATATATTAACCAAAATCAAAATAAAAAACAACAAATATATTAACATACT	6
QY	1520	CTGCACCTTTAAATGGCTTTCTTTATGATCTTTACATTTATTTAGATAGTAGTGAAT	6
Db	6712	CTTAAACAAATTAATATATTTTAAAAAACCACAAAAATATTAATATTAATCTAAAAA	6
QY	1580	TTAATAGACCACTTACATTAACATTAATTTATTTAGATAGTAGTGAATGAT	6
Db	6652	TATCTTTAAAAAATATATATTTTAAAAAATAAAAAATCTAACAA - AAAAAA	6
QY	1640	CCAAATACATCTTTGGATGTTAAACTAATCTGTTCTTCCTGACGGTAAATATTA	6
Db	6594	AAAAAACAACAAATTAATATCTAATTAATTAACAAATTAATTAACAAACCAATTA	6
QY	1700	ATCATCGAGTAAAAAAGTTGGTGTCTATTTCCGATGCATGAAGATTAACCTATG	6
Db	6534	AATATATCTTTTAAATTTTCAATATCTAAACAAACCTCATCACTAAAAAATAAT	6
QY	1760	ACTTTAATTTTGAANAATGAACCTTTTACTCATATTAATTAATTAACGATATGTT	6
Db	6474	TTTCAATTTAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	1820	TGCATATACACCCCTACAAACGTGTATCTCAATTTTTCGCAATTAATTAATTA	6
Db	6414	TAAATCATTATTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	1880	GAATCAATGCT - - - ACTATCAATAGAAAGACGTGATTAATTAATTTAATTA	6
Db	6354	AAAAAATCTTTAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	1936	GACAAATTTTGAANAATGATTAATTTTCAACAAATTAATTAATTAATTAATTA	6
Db	6294	TACATATTTTCACTCATATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	1996	ATGATTTTCCATGTTCTTAAATTAATTTTATTTATTTAGTTAATTAATTAATTA	6
Db	6234	ATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	2056	CCAAATATAGTTGGAATTCAAATTAATTAATTAATTTTTCGAAATTCACAAATTA	6
Db	6174	CTCTTAAACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	2116	TAAATTTAGCATTAACATGATGAAAGTCCAAAAAATTTTGTAAACGAAC	6
Db	6114	AATAAAAATCAAAAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	2176	TCCAATTTTATTTTATGAGAACAAAGATTAACAGATTAATTAATTTTGTGTGA	6
Db	6054	TCACCATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	2236	ATGACATGATTAATTAATTAAGCAATTTTAAAAAATTTATTAATTAATTAATTA	6
Db	5994	ATTTTCAACATTAATTAACAACTTACTTAATAAAATTTATTTCTTAATTAATTA	6
QY	2296	AAAGTATGTTATCTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
Db	5934	AAATCCACGTTTATTAATCTTCTCAAAATCTCATTAATTAATTAATTAATTAATTA	6
QY	2356	CAATGAAGGAAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTA	6
Db	5874	TATACAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6
QY	2416	TTTGACTTTGA 2428	5815
Db	5814	CCCTATCTTTTCA 5802	

Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;

Query Match 2.5%; Score 81.6; DB 24; Length 5689;
Best Local Similarity 49.1%; Pred. No. 0.00044;
Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;

QY 1824 ATATGACACCCCTCTACACTGTGATGTCATTTTTCGCAATTTTAATAGCAAT 1883
DB 2110 AAAATCGCAATCGGCAACCGGCAATATCAAAAAATTTCTCCGAAAAA 2051
QY 1884 TCAATGCTACTATCAATAGAAACAGTGTATATACATTTTAATTAAGCAAAAT 1943
DB 2050 TATAAAAAACAATAAAAAATTTTATTTTATATTTTATTTTATTTT 1991
QY 1944 TTTTGAATAATGTTATATTTCTACAAATTTATTAATATGATGCTTATATTT 2003
DB 1990 ATATTTTAAAAATATTAATAAATAAATAAATAAATAAATAAATAAATAA 1931
QY 2004 CCAATGCTCTTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2063
DB 1930 NATAAAAAATAAT 1871
QY 2064 AGTTGGT-GAATCAATATCTCATTAATTTTGAATCTCAAAATTTATATAT 2122
DB 1870 AATTAAT 1811
QY 2123 TACTCAATACAAATGCAT-AGAAAGTTCAAAAAATTTTGAACAGAACTTCGAA 2181
DB 1810 TATTAAT 1751
QY 2182 TTTTATTTTATGACAGAAATATACAGATAGAAATTTTGTGTGATGGA 2241
DB 1750 TAAAAAT 1691
QY 2242 GTAGTAT 2301
DB 1690 TACCAAT 1631
QY 2302 TGTATCTAGTAGGTATATATATATATATATATATATATATATATATATAT 2360
DB 1630 TATTAAT 1571
QY 2361 AAAACGAAT 2400
DB 1570 ATAAAAAAT 1531

RESULT 12

ABL33054
ID ABL33054 standard; DNA; 6294 BP.

AC ABL33054;
DT 26-MAR-2002 (first entry)
XX

DE Human immune system associated gene SEQ ID NO: 1027.

KW Human; immune system disease; cytosine methylation; antiaesthetic;
KW antiarteriosclerotic; antianemic; cytosine; neotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineurotic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX Olek A. Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
XX

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation

PS Claim 1; SEQ ID NO 1027; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 6294 BP; 1870 A; 110 C; 1224 G; 3090 T; 0 other;

Query Match 2.5%; Score 80.6; DB 24; Length 6294;
Best Local Similarity 45.4%; Pred. No. 0.00063;
Matches 410; Conservative 0; Mismatches 484; Indels 9; Gaps 3;

QY 1518 ATCTGCACTATTAATGCTCTTCTTAATGATCTTTACATATTTATTTAGTGA 1577
DB 2352 ATATTAATTAATAAATATTTTGTGTATATATATATATATATATATATATAT 2411
QY 1578 ATTAATTAAGGCACTGTATACATATATATATATATATATATATATATATAT 1637
DB 2412 GAATATTAAGATATGTTTAAATATTTTAAATATTTTAAATGATATATATAT 2471
QY 1638 TTCCAAATACATCTTGTGATGTTTAAATCTTATCTTCTCCGATATATAT 1697
DB 2472 TTTTATTTGATTTTATATATATATATATATATATATATATATATATATAT 2531
QY 1698 TATCATCGAGTAAAAAAGCTTGTCTATATTTCCGATGATGAGATATATAT 1757
DB 2532 TTAATAATTTATTTATATATATATATATATATATATATATATATATATAT 2591
QY 1758 TGACTTAATTTTGAATATGTAACCTTTTACTCATATATATATATATATAT 1817
DB 2592 TTTGTTAGTTTATGAGTTTATGAG---TAAATATATATATATATATATATAT 2647
QY 1818 GTTGCCATATGAGGCTCTACACCTGATGATGATGATGATGATGATGATGAT 1877
DB 2648 AATTAATTAAT 2707
QY 1878 AGCAATTCAGTCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1937
DB 2708 ATAAAAAT 2763
QY 1938 CAAATTTTGAATAATGATATATTTCTACATATATATATATATATATATATAT 1997
DB 2764 TGGATTTTATTAAGGATATATATATATATATATATATATATATATATATAT 2833
QY 1998 GATTTTCTATGTTCTTAATATATATATATATATATATATATATATATATAT 2057
DB 2824 ATGTTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2883
QY 2058 AATTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2117
DB 2884 TTTTATTAAGAAATTTTATATGTTGGGATGATATTTGATATATATATATAT 2943
QY 2118 ATATTT-TAGTCAATATACATGATGATGATGATGATGATGATGATGATGAT 2176

ABA92787
 ID ABA92787 standard; DNA: 640681 BP.
 AC ABA92787;
 XX
 DT 27-MAR-2002 (first entry)
 DE Buchnera sp. genomic DNA SEQ ID NO:1.
 XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
 KM circular; ds.
 XX
 OS Buchnera sp.
 XX JP2001292771-A.
 XX
 PD 23-OCT-2001.
 XX
 PF 07-APR-2000; 2000JP-0107160.
 XX
 PR 07-APR-2000; 2000JP-0107160.
 XX
 PA (RIKA) RIKGAKU KENKYUSHO.
 XX
 DR WPI: 2002-126043/17.
 XX
 PT A genomic DNA of cockroach-symbiotic bacterium
 XX
 PS Claim 1; Page 16-230; 237pp; Japanese.
 XX
 CC The present invention describes a gene (I) derived from Buchnera sp.
 CC containing the DNA (a) or (b), (a) has a fully defined base pair
 CC sequence selected from a table of sequences found in the Buchnera sp.
 CC genomic DNA of ABA92787 given in the specification or is a DNA selected
 CC from complementary DNA sequences, and (b) is a DNA which hybridises with
 CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
 CC vector (II) containing (I); (2) a transformant (III) containing (II);
 CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
 CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (C) or
 CC (d), (C) is a DNA containing a fully defined sequence given in ABA92788
 CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
 CC method for the preparation of a protein in which (III) is cultured and
 CC the expression protein of the objective protein is collected from the
 CC resultant culture. The DNA is useful for developing agricultural
 CC chemicals for exterminating cockroaches. The present sequence represents
 CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
 CC present invention.
 CC
 XX
 XX
 SO Sequence 640681 BP: 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
 Query Match 2.5%; Score 80; DB 24; Length 640681;
 Best Local Similarity 45.8%; Pred. No. 0.00086;
 Matches 395; Conservative 0; Mismatches 455; Indels 12; Gaps 3;
 QY 1559 ATATTATTTAGTTGGAATTTAATGAGCGCACTTGTAACATTTATTTATTT 1618
 DB 324686 ATATTGTAATTTCTTTTAAATACCAATATTTATTCATCAATTAATATGCGCAC 324745
 QY 1619 AGATACAGTATGATTTATTCACAAATACATACATTTGATTTGATTTATTTGTTT 1678
 DB 324746 AATCTTTTGTACAAAGAAATTAATTTTATTTTATCTAATTAATTCATTTTGTATC 324805
 QY 1679 CTTCCTACGCTATTAATTAATTAATCAGAGTAATAAAGTTTGTCTTAATTTTCGGAT 1738
 DB 324806 TCTTCTATCATGTAAATAATCAATTTCTTCTTAATCAATTAATTTCTTAATTAAT 324865
 QY 1739 GCATGAAGTAACCAATGACTTTATTTTGAATAATGTAACCTTTTACTCATAGA 1798
 DB 324866 TGTGGGTGTAATTTATGCTCATTAATAAATCTTAATAATTTTAAAGATTTTCTCTC 324925
 QY 1799 TTATATTACGCTATTTTGTGTCATTAATGACAGCCTTCATACAGTGTATGCAATTT 1858
 DB 324926 GTCAAGAAATATTTATTTTCTGAAATGAATTAATTAATGACATTAATCTGATTTT 324985

QY 1859 TTTCGCAATATTAAATTTAGGAATTCATGCTACTATCAATAGAGAAGACAGTGA 1918
 DB 324986 TTTTTCGATTAATTACATTTAATAAATGATCTGATTAATTTTGATATTTAAATTA 325045
 QY 1919 TTACATTTTAATTTAAGACAAATTTTGAATAATGTTATTAATTTCTACATATTTAT 1978
 DB 325046 ATTTTGTGTTTAAATAATTTTATTTTGTGAATTTTCAATGAAACGATTTCTTAAAT 325105
 QY 1979 AAAATGATGCTTAATGATTTGCTTAATTTTCTTAATTTTATTTTATTTTATTTAT 2033
 DB 325106 AAAGTAAATTTTATTTTAATTTAATTAATTAATTAATTTTATTTATTTATTTA 325165
 QY 2034 TTAGTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2093
 DB 325166 ATTTAAATAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 325225
 QY 2094 TTTTGTGAATCTCAAAAT-----TATTAATTTTATGCAATTAATGATGATGAACTT 2148
 DB 325226 ATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 325285
 QY 2149 CCAAAAAAATTTGTTTACGAAACCTTCAAAATTTTATTTTATTTGGAACAAGAAATA 2208
 DB 325286 ATTAATAATTAATTTATTAATTAATTAATTAATTAATTAATTTTCTTTTATTTTAAAT 325345
 QY 2209 ACAGATAGAAACTATTTTGTGGAATGGAAGTAAATTAATTAATTAATTAATTAATTA 2268
 DB 325346 ACTATTAATGATA--ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTA 325403
 QY 2269 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2328
 DB 325404 TCAATTAATAATTAATTAATTAATTAATTAATTAATTTCTTTTATTTTATTTA 325463
 QY 2229 GCATGTGCGCATTCGAAATTTGGAACAATGAAACGAATTAATTAATTAATTAATTA 2388
 DB 325464 GAATTTGATGCAATTCGATTAATAATTAATTTTATTTTCAATTTGATTAATTTCA 325523
 QY 2389 ATTAATAATAATTTGATTAAT 2410
 DB 325524 TTCAATTAATAATTTTATTAATTT 325545
 RESULT 15
 ABR31511/c
 ID ABR31511 standard; DNA: 47108 BP.
 XX
 AC ABR31511;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified complementary DNA #177.
 XX
 KW Human; signal transduction associated gene; cytosine methylation strate;
 KW Cpg island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytosolic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 PD WO200200926-A2.
 PD 03-JAN-2002.
 PD 29-JUN-2001; 2001WO-EP07472.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-147896/19.

The present invention relates to chemically modified DNA sequences of signal transduction associated genes. The DNA sequences are chemically modified using a solution of bisulphite, hydrogen sulphite or disulphite. Also disclosed are oligonucleotides and/or PNA oligomers for detecting the cytosine methylation state (CpG islands) of these genes, and a method for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with signal transduction. The genomic DNA can be obtained from cells or cellular components which contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, and cerebrospinal fluid, tissue embedded in paraffin such as tissue from eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver, histologic object slides, and all their possible combinations. The sequences of the invention are useful for the diagnosis and therapy of diseases associated with signal transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA sequences of different genes associated with signal transduction, or their complementary sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

sequence 47108 BP; 15349 A; 429 C; 8736 G; 22594 T; 0 other;

Query Match	2.5%;	Score 79.6;	DB 24;	Length 47108;
Best Local Similarity	44.48%;	Pred. No. 0	00004;	

Matches	457;	Conservative	0;	Mismatches	564;	Indels	9;	Gaps	3;
---------	------	--------------	----	------------	------	--------	----	------	----

1074 TTTCAAGCTATTATTCAAAGCTTACAAACTGAATATAGCTTGAATCCTTTAGAAAAATTTT 1133

1033	TTAAATCCTTTTACCAAAAAAAGCATATTAATTAATTACCAACTTAATAAATCT	974
1134	CGATTTTCTTTTACCAAAAAAAGCATATTAATTAATTACCAACTTAATAAATCT	

1134 GGAATTTACCGGTTGTTATGTAATATATGATTAGTGGTAAACAATAATGTTATCATTTA 1193

1194 GCGCTCAACCTATGATTTCTGGCTTTT
213 GAGGACCTAAGAAAAATATCAAAATTCATTCTTAATTATTAATATATATATA 914

913 ATTATTAATAAATAATGCTGCTTACAGAAAAACAACCTTAAGAGAACTTAACATATC 1253

1254 CATATATGGGTATGCTAATTACCAAAATTTAACAAAT 854

[illegible][illegible]

793 AAATTTTAAATTAAACACAGACACACATTAATTCCTAATAATGCGACGTATGTTACC 1370

1371 ATTCACGCAATATCATCTTGTGGTTGGTCAACAATAAAGCAAAACAAATTAAC- 735

b

734 TTTCTTAATTCACAACTAPCATAAAAATTCTTACTGATAAAGCAGAATGGCCCGGTGC
| | | | |
|| |||||||||CAAGAAAAAAAAAGGA 1430

1431 AACGATTTTCTGGATTCATTCATCATGATCTAAATGATCATAGATCTTTTTCCCTTTTCTCT 1460

674 AAATCCATATAAAATACACACAATTTTCTMAAATAATATTTCCTTAAGAACGCT 615

1491 TTGGAAGTCTCTACACAGCGTGTACCACTGCAACTATTAAATTGCTTTCTCTAAAGCGA 1500

614 TCCGAATAACCAACACTTTAAATACCAAAATCTTTTAACTATATTATATATAT 555

1551 TCTTACATATTATTGTAGTTCGATTATTAAGAGCGAAGCTGTGACATTACATA 1610

554 ATATACTTAACTACCTTTTAAACAATAATACACATAATTAAATTATCTCTTAAATAC 495

1611 TTTATATACATACTAGTATGATTAATCCAAAT-----ACATACTTGGATGTTTAA 1665

494 TAACTACTCCAAATTAAATAATATTTTACTATATTTCTAAATAATATCTAAATCTAAA 435

QY 1666 CTTAATCTGTTCTCTCTACGGTATAATATTAATCATCGAGCTAATAAAAAAGTTTTC 1723
 Db 434 TTCACCTTTAACCTTAAAAATTTACTTTATTTTATTTCTCTCCTCATATCTAAATTT 375

QY 1726 TTATTTTCGCATGCATGAAGCATAACTTAATGACTTTAATTTTTTGAANAATGTACCC 1785

Db
374 TCATTTCTACTAATATCATATATATATTTAATAATTACCAATTCCTCCCTATATAT 315

QY 1786 TTTTACTCATAGATTAAATTAACCGTATGTTTTCGTCCATATGACAGCCTCTACAACTG 1845

Db 314 TAATTCCTAATCTACTATTAATCTTAATAAACAAATATATATTTTCTTTCCTTTTCAAAAAATA 255

1846 TGATAGTCAATTTTCTCGCAATATTAAATTAGGAATTCATGCTACTATCAATAGAAG 1905

254 TTATATTATAAACTATATAAAAAAAAAATTTCATTAATAAAAAAAAACTTTATTATTAATAAT 195

1906 A A A C A G C T G A G T A T T A C A T T T A A T T T A A G C A A A A T T T T G A A A A A G T T A T A T T T C 1965

134 ATATACACCAATACCCCTTCAATTAATTCAAATATATCTAAATTCAAAAAATT 135

1508 AACCAAAATATTAATAAAGATGCCATATATGTATTTCCCTATGTCTTAAATAATTTT 2025

107 C TGGGAGTAAAGAAATTCCTATACATAAATATATAACCTCTATCATTTTACCAACTTC 75

Dd
Db

74 CTTAAATACCAATTATAAATTAAATTAATGTTGGTGAATTCATAATATCTC 2085

-----CAGGCGAATTTTAA 15
2086 CATTATATTT 2095

Db 14 TTTTAAATTT 5

Search completed: June 10, 2003, 22:22:47
Job time : 737.005 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 152.509 Seconds

(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtgggtattatattgttgc.....gagagagagaactactctcc 3202

Scoring table: IDENTITY-NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	ID	Description
No.						
1	77.4	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
2	68.6	2.1	837	4	US-08-998-416-288	Sequence 288, App
3	67.4	2.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
4	65.8	2.1	615	4	US-08-998-416-186	Sequence 186, App
5	65.4	2.0	636	4	US-08-998-416-1137	Sequence 1137, App
6	65.4	2.0	20674	4	US-09-641-638-651	Sequence 651, App
7	63.6	2.0	8920	2	US-08-446-855A-1	Sequence 1, Appl
8	63.6	2.0	20674	4	US-09-150-741-1	Sequence 1, Appl
9	63	2.0	20674	4	US-09-641-638-651	Sequence 651, App
10	62.6	2.0	4818	3	US-08-817-926-27	Sequence 27, Appl
11	61.8	1.9	5852	1	US-07-867-106-2	Sequence 2, Appl
12	61.2	1.9	1431	4	US-09-316-083-2	Sequence 2, Appl
13	60.8	1.9	4818	3	US-08-817-926-27	Sequence 27, Appl
14	60.6	1.9	837	4	US-08-998-416-288	Sequence 288, App
15	60.4	1.9	8920	2	US-08-446-855A-1	Sequence 1, Appl
16	60.4	1.9	8920	2	US-09-150-741-1	Sequence 1, Appl
17	60.2	1.9	665	2	US-08-883-795A-36	Sequence 36, Appl
18	60	1.9	1864	4	US-09-468-265-4	Sequence 4, Appl
19	59.8	1.9	6124	4	US-08-213-419B-3	Sequence 3, Appl
20	59.2	1.8	51952	3	US-08-947-823-1	Sequence 535, App
21	59	1.8	827	4	US-08-998-416-535	Sequence 535, App
22	58.4	1.8	665	2	US-08-883-795A-36	Sequence 36, App
23	58.2	1.8	12793	4	US-09-004-838-124	Sequence 124, App
24	58.2	1.8	6124	4	US-08-213-419B-3	Sequence 3, Appl
25	57.8	1.8	1511	1	US-07-991-867B-8	Sequence 8, Appl
26	57.8	1.8	1511	1	US-08-107-755A-8	Sequence 8, Appl
27	57.8	1.8	1511	2	US-08-544-332-8	Sequence 8, Appl

c	28	57.8	1.8	1511	4	US-09-370-861A-8	Sequence 8, Appl
	29	57.4	1.8	3095	6	5231168-1	Patent No. 5231168
	30	56.6	1.8	168575	4	US-09-426-290-1	Sequence 1, Appl
	31	56.4	1.8	2058	2	US-08-749-391-1	Sequence 1, Appl
	32	56.4	1.8	2058	2	US-09-390-200-1	Sequence 1, Appl
	33	56.4	1.8	4673	3	US-07-638-431-1	Sequence 1, Appl
	34	56.4	1.8	4673	5	PCT-US92-00018-1	Sequence 1, Appl
	35	56.2	1.8	1431	4	US-09-316-083-2	Sequence 2, Appl
	36	56.2	1.8	2960	3	US-08-913-842-3	Sequence 3, Appl
	37	55.6	1.7	701	4	US-08-998-416-701	Sequence 701, App
	38	55.6	1.7	5852	1	US-07-867-106-2	Sequence 2, Appl
	39	55.2	1.7	1406	3	US-08-913-842-6	Sequence 6, Appl
	40	55.2	1.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
	41	55.2	1.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
	42	55.2	1.7	1511	2	US-08-544-332-8	Sequence 8, Appl
	43	55.2	1.7	1511	4	US-09-370-861A-8	Sequence 8, Appl
	44	55	1.7	1850	3	US-08-617-860B-32	Sequence 32, Appl
	45	55	1.7	4098	2	US-08-605-106-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 593827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Williams, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
Query Match 2.4%; Score 77.4; DB 2; Length 19124;
Best Local Similarity 42.6%; Pred. No. 1.8e-06;

Matches 590; Conservative 0; Mismatches 781; Indels 14; Gaps 3;

```

QY 1013 TTAATAATGATTTTACAGCTCACTAGTGTAGATTACTAGCATTAAGCATTAAGAGCC 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 TTATATATACATATATATATAGTTCCTATTAATAATTAATTAATTAATTAATTAAT 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1073 GTTCAAGCTATTTTACAAAGTTACAACTGAATATAGCTTGAATCTTTAGAAAATTT 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TTCCTGTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1133 TGAATTAACCGGTGTTATGTAATAATTAATTAATTAATTAATTAATTAATTAAT 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AATTTCTTAATTTCTTTTATGCAACAAAAACATTAAGTAATTTCTACTATACACAAAA 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1193 AGTGTGACATATATACATTAATTTCTTACAGAAAAACAACTAAGAGAGTAAACATAT 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 AAAAAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1253 CCATATATGGGTATGCTATACCTTTACAGTATGCTATATCTAGAGACTAAGAAATGTTAT 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 TCACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1313 GTGATGTGATAAATGAATAATCACACGGGTGTAATATTAATTAATTAATTAATTAAT 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TTTAACATAGAAAGTTAAGAAATACATTTTATTTTATTTTATTAATTAATTAATTAAT 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1373 CACTGCAAAATATCATCTTGTGTCGCAACATTAATAAACAACAAAAACAAAAAGAAAA 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 ATATATATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1433 CGATTTTCTTGATTTCCATTCATGATCTAATATGCTATAGATCTTTGGGTACAGTTT 1492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 TCTTTTATC-----ACATGGAATATTAATAAATTAATTTGATTTTATGATATATAT 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1493 CGAAGTCTCTCAAGCGGTGTAACCATCTGCACTAATTAATTAATTAATTAATTAAT 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 ATGCGTTATATATATATATAGTCTTTATTAATTAATTAATTAATTAATTAATTAAT 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1553 TTTAACATATTAATTTAGTTGAATTAATAAGGCACTGTAACTTACATTAATTAAT 1612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1613 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 TTATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1673 TTGTTCTCTCAAGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GCATATACATGACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1733 CGCGATGCAATGAAGATTAACCTTAATTAATTTTGAATAATGTAACCTTTTACT 1792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 AATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1793 CATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1853 CAATTTTCTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 CTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1913 TGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1973 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 TAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2033 TTTAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1238 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 2093 ATTTTGAATTCACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1295 TTGTATATACATTAACAGCTAAGAAACCTATACAAATCTGATATCTAATTAATTAAT 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2153 AAAAAATTTTGTACAGAAACCTTCCAAATTTTATTTTATTTATGACACAGAAATACAG 2212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1355 ATATATCTTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2213 ATAGAAATCTATTTTGTGTGAATGAAGTAGTAATTAATTAATTAATTAATTAATTAAT 2272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1415 TAAATATTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2273 ATATATAGCTATATACGCTCAAAAGTATGTTATCTAGTAGTGTAAATTAATTAATGAT 2332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1475 AAAACTTCAAAACATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2333 GGTGATTCAGAAATGGGACACACATATAAACGAAATTAATTAATTAATTAATTAAT 2392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1535 AGAGAAACGTAGAACATACCAAAAAAATTAAGAACAAAAAAGAAATTAATTAATTAAT 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2393 ATAAA 2397
    |||
Db 1595 TAAAA 1599

```

RESULT 2
US-08-998-416-288
Sequence 288, Application US/08998416

Patent No. 6239264
GENERAL INFORMATION:

APPLICANT: Phillipson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwalls Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 288:

SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

QY 1137 ATACCGGTTGTTATGTAATAGATTGTCGTAACAAATATGTTAAATCAATTAAGT 1196
DB 6327 AAATAGTAAAAATACACATTAATAATATATATCAAAATATGAGTTAATAATAAAT 6268
QY 1197 GTCACATATACATTAATCCCTTACAGAAAAAACAACCTTAAGAACTTAACATTAACAT 1256
DB 6267 TTTATGTCCTAT 6208
QY 1257 ATATGGTATGCTATACCTTTCAC -GTATGCTATAGTACAGACTTAAGAACTTAAGT 1315
DB 6207 TACTAT 6148
QY 1316 ATGTCAT 1375
DB 6147 AT 6088
QY 1376 TGCAT 1435
DB 6087 TAT 6028
QY 1436 TTTTCTGATTCAT 1495
DB 6027 TTTAT 5968
QY 1496 AGTCCTCTACACCGGTAAACCATCTG -CACTATATATATATATATATATATATAT 1553
DB 5967 AT 5908
QY 1554 TTAAT 1613
DB 5907 TAT 5848
QY 1614 AT 1673
DB 5847 AGTAT 5788
QY 1674 TGTTCCTCTACCGGTAAACCATCTG -CACTATATATATATATATATATATATAT 1733
DB 5787 GAT 5728
QY 1734 GCGAT 1793
DB 5727 TGTAT 5669
QY 1794 AT 1853
DB 5668 CCAATCAAAAAACAT 5609
QY 1854 AATTTTCTGCAAAAT 1913
DB 5608 TATTTTGTATGAATATCTATATATATATATATATATATATATATATATATATAT 5559
QY 1914 GAGTAT 1973
DB 5558 ---TATTTAAATTTTTCAAAAAATCAATGATTAATATATATATATATATATATAT 5502
QY 1974 TATTTAAAT 2033
DB 5501 TGCAT 5442
QY 2034 TTAGTAT 2090
DB 5441 TTTCTAT 5382
QY 2091 ---ATATTTTGAAT 2147
DB 5381 GACTTAT 5322
QY 2148 TCCAAAAAATTTTGTATACAGAAATCCCAATTTTGTATATATATATATATATATATAT 2207
DB 5321 TATTTACTTAT 5262
QY 2208 AACAGAT 2267

DB 5261 TAAGCTATGGAACACTACAAATCAGATATATATATATATATATATATATATATAT 5202
QY 2268 AAAAAT 2327
DB 5201 AATTTAT 5142
QY 2328 TGCATGATGATTCAGATAT 2348
DB 5141 AAAAACAACAATAATCGAAT 5121

RESULT 4.
US-08-998-416-186
Sequence 186, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIT
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 36,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1074RP
US-08-998-416-186

Query Match 218; Score 65.8; DB 4; Length 615;
Best Local Similarity 48.38; Pred. No. 0.00015;
Matches 216; Conservative 0; Mismatches 227; Indels 4; Gaps 1;

QY 1750 AACCTAT 1809
DB 56 ATATTTAT 115
QY 1810 ATGTTTGTGTCAT 1869


```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAGI692RP
US-08-998-416-1137

Query Match          2.0%   Score 65.4; DB 4; Length 636;
Best Local Similarity 47.1%; Pred. No. 0.00018;
Matches 269; Conservative 0; Mismatches 296; Indels 6; Gaps 2;

QY      1841 AACTGATAGTCACATTTTCTGCAAATTATTAATTGGAATCAATGCTACTCATCAT 1900
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       29  AAGATTAAATAAACCTTTTTATTATATATATTAGTGTTAAATATTATTAACGTATT 88

QY      1901 AGAAGAAACAGCTGAGTATTCATTTAATTAAAGACAAAATTTTGAAAAATGTTATA 1960
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       89  ATCATTCTTTTAATTAATTAATTAATTTGGATTATTAATACCTTTTATATATTATTAATTA 148

QY      1961 ATTCTCAACAATTAATTAATAAATATGATGCCATAATGTAATTCCTATGTCCTAAAT 2015
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       149 TTTACTTAATTCATCATTTATTAAATTTATATTAATTAATAAATAATTAATTAATGTAAT 208

QY      2020 ATTTTCTTTTATATTGTGTTAAATAACATTTATGAACCAATAATAGTTGGTAATTCAAA 2075
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       209 ACATTTAGTCTGATAGTCCAAATTTTAAATTTGTTATTAATAATTAATTAATGATTAATAT 258

```

RESULT 6
 US-09-641-638-651
 ; Sequence 651, Application US/09641638
 ; Patent No. 6432648
 ; GENERAL INFORMATION:
 ; APPLICANT: Blumenfeld, Marta
 ; APPLICANT: Bougueleret, Lydie
 ; APPLICANT: Chumakov, Ilya
 ; APPLICANT: Cohen, Annick
 ; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
 ; FILE REFERENCE: GENSET.051CPI
 ; CURRENT APPLICATION NUMBER: US/09/641,638
 ; PRIOR FILING DATE: 2000-08-16
 ; PRIOR APPLICATION NUMBER: US 09/502,330
 ; PRIOR FILING DATE: 2000-02-11
 ; PRIOR APPLICATION NUMBER: US 60/133,200
 ; PRIOR FILING DATE: 1999-05-07
 ; PRIOR APPLICATION NUMBER: US 09/275,267
 ; PRIOR FILING DATE: 1999-03-23
 ; PRIOR APPLICATION NUMBER: US 60/119,917

PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5936..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTT
NAME/KEY: allele
LOCATION: 2048

OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G

```

: NAME/KEY: allele
: LOCATION: 8608
: OTHER INFORMATION: 10-349-47 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 8658
: OTHER INFORMATION: 10-349-97 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 8703
: OTHER INFORMATION: 10-349-142 : polymorphic base G or C
: NAME/KEY: allele
: LOCATION: 8777
: OTHER INFORMATION: 10-349-216 : deletion of CTG
: NAME/KEY: allele
: LOCATION: 8785
: OTHER INFORMATION: 10-349-224 : polymorphic base G or T
: NAME/KEY: allele
: LOCATION: 8926
: OTHER INFORMATION: 10-349-368 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 12171
: OTHER INFORMATION: 10-350-72 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 12429
: OTHER INFORMATION: 10-350-332 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 13341
: OTHER INFORMATION: 10-507-170 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 13492
: OTHER INFORMATION: 10-507-321 : polymorphic base A or C
: NAME/KEY: allele
: LOCATION: 13524
: OTHER INFORMATION: 10-507-353 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 13535

```

```

Query Match      2.0%; Score 65.4; DB 4; Length 20674;
Best Local Similarity 52.6%; Pred. No. 0.0004;
Matches 235; Conservative 0; Mismatches 206; Indels 6; Gaps 4;

```

```

QY 1842 ACTGATGATGCTATTTTTCGCAATATTAATAGGAATCAATGCTACTATACATA 1901
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11120 AATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1902 GAAGAACAGCGATATACCTTTAATTAATTAAGCAAAATTTTGAATAATGTTATA 1961
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11179 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1962 TTCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2021
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11239 TTTTAAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2022 TTTTAAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2081
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11299 TTTTAAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11357
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2082 TCTCATTAAT---ATTTTGAATCTACAAATTAATTAATTAATTAATTAATTAATTA 2138
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11358 TTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2139 ATAGAAAGTCCAAAAAATTT-TGTTAAACGAAGCTCAAAATTTTAAATTTAAATG 2197
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11418 CTTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2198 AACCAAGAAATACAGATAGCAATTTTGTGGAATGAGAGTAAATATACATTA 2257
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11478 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 11537
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2258 AGCAATTTTAAATTAATTAATTAAGC 2284
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11538 AAAATTAATTTTAATTAATTAATTAAGAC 11564
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-08-446-855A-1
: Sequence 1, Application US/08446855A
: Patent No. 5849573
: GENERAL INFORMATION:
: APPLICANT: Stewart, Thomas S
: APPLICANT: Flores, Maria V
: APPLICANT: O'Sullivan, William J
: TITLE OF INVENTION: Nucleotide sequence encoding carboxymyl
: NUMBER OF INVENTIONS: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon & Vanderhye PC
: STREET: 1100 No. 5849573th Glebe Road, 8th Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22201-4714
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,855A
: FILING DATE: 06-Jul-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mitchard, Leonard C
: REGISTRATION NUMBER: 29,009
: REFERENCE/DOCKET NUMBER: 47-80
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-816-4000
: TELEFAX: 703-816-4100
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic
: US-08-446-855A-1

```

```

Query Match      2.0%; Score 63.6; DB 2; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00074;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

```

```

QY 1874 AATTGAATTCATGCTACTATCAATAGAGAAACGCTGATTAATTAATTAATTA 1933
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 152 AAAAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1934 AAGACAAATTTTGAATAATGTTATA-ATTTCTAACAAATTAATTAATTAATTAATTA 1992
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 212 ATATATATGATTTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 271
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1993 ATAAATGATTTCTATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2052
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 272 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 331
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2053 GAACCAATATAGTTGGTGAATTCAAATATCTCCATTAATTTTGAATCTACAAAT 2112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 332 TATATATATATATATATATGTTGTGTCATTTGTTTAATAATTAATTAATTAATTA 391
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2113 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 392 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 451
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2173 ACTTCAATTTTATTTTATTTTATGCA-----ACAGAAATTAACAGATGAAGAC 2221
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 ATATAGTTTAACTCTTATTAATAAACAATCCTATATATTAATTAATTAATTAATTA 511
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2222 TATTTTGTGGAATGAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2281
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Query Match	2.0%	Score 63.6	DB 4	Length 8920
Best Local Similarity	46.8%	Pred. No. 0.00074		
Matches 278	Conservative 0	Mismatches 304	Indels 12	Gaps 2
QY	1874	AATTGAAGATTCAGTCTACTATCAATAGAGAAAACAGCTAGATTTACATTTTAATT	1933	
DB	152	AAAAAAAAATATAAAAAAAAAAAGATATAAAATATTTATTTGATATAGATAT	211	
QY	1934	AAGCAAAATTTTGGAAAAATGTTATAT-ATTTCATACAAATTTATATAAATATGTCCT	1999	
DB	212	ATATATAAGATTAATCATTTATATATACATAGATATAAAAAACCTTTTTTTTTTTTTCT	271	
QY	1993	ATAATGATTTCCATGCTCTTAAATATTTTTTTTATTTAGTTATTAATACATAT	2052	
DB	272	TTATATTTATTAACAATACATTTAAAGTATTTTATATATATATATATATATATATA	331	
QY	2053	GAACCAATATAGTGGTGAATTAATATCTCATTAATATTTTTTGAATCTTCAAT	2112	
DB	332	TATATATATATATATGTTGTGTCTCATTTGTTTAAATTTACTGAAATATATAA	391	
QY	2113	TATTAATATTTAGCATATACAAAGCATAGAAAGCTCCAAAAAATTTGTTAAACAA	2172	
DB	392	TATTAATATATTTCCATTTATATATATATATATATATATATATATATATATAT	451	
QY	2173	ACTTCAAAATTTTTTTTTTTATGA-----ACAGAAATACACAGATGAAAC	2221	
DB	452	ATATAGTTTACACCTCTTATATAAATAAACCATCCTATATATATTATACAAATATATA	511	
QY	2222	TATTTGTTGGGAATGGAGTAGTATATATACATTAAGCAATTTTAAAAATATATATA	2281	
DB	512	TCCCCAATATTTGGTGCTTATATATTTATATATATATATATATATATATATATAT	571	

```

? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1123..3123
? OTHER INFORMATION: 5' regulatory region
? NAME/KEY: exon
? LOCATION: 3124..3297
? OTHER INFORMATION: exon 1
? NAME/KEY: exon
? LOCATION: 3871..4072
? OTHER INFORMATION: exon 2
? NAME/KEY: exon
? LOCATION: 5552..5633
? OTHER INFORMATION: exon 3
? NAME/KEY: exon
? LOCATION: 5758..5880
? OTHER INFORMATION: exon 4
? NAME/KEY: exon
? LOCATION: 5996..6099
? OTHER INFORMATION: exon 5
? NAME/KEY: exon
? LOCATION: 6349..6509
? OTHER INFORMATION: exon 6
? NAME/KEY: exon
? LOCATION: 7379..7522
? OTHER INFORMATION: exon 7
? NAME/KEY: exon
? LOCATION: 8645..8854
? OTHER INFORMATION: exon 8
? NAME/KEY: exon
? LOCATION: 12254..12340
? OTHER INFORMATION: exon 9
? NAME/KEY: exon

```

LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc-feature
LOCATION: 17555..20674
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903

OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 2.0%; Score 63; DB 4; Length 20674;
Best Local Similarity 48.0%; Pred. No. 0.0012;
Matches 224; Conservative 0; Mismatches 230; Indels 13; Gaps 1;

QY 1825 TAATGACAGCCTCACAAGCTGATAGCAATTTTCTGCAATATTAATTGGAATT 1884
DB 11554 TAAATTTAAATATTTTCTAGCTAATTAATTAATTAATTAATTAATTAAT 11495
QY 1885 CAATGCTACTATCAATGAGAAACAGCTGATTCATTTTAAATTAAGACAAATT 1944
DB 11494 ATTAATATTTAATTTAATTAATTAATTTAATTAATTAATTAATTAATTAAT 11435
QY 1945 TTGAAAAAGTGTATTAATTTCTAACAATATTAATTAATTAATTAATTAATTTTC 2004
DB 11434 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11375
QY 2005 CTATGCTCTTAAATATTTTATTAATTTAGTTAATTAATTAATTAATTAATTAAT 2064
DB 11374 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11315
QY 2065 GTTGGAATTCATTAATTC-----TCCATTAATTAATTTTGAAGCTACAAA 2111
DB 11314 TTAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11255
QY 2112 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11254
DB 11254 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11195
QY 2172 AACTCCAAATTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2231
DB 11194 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11135
QY 2232 TCGAATGGAAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2278
DB 11134 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11088

RESULT 10
US-08-817-926-27
Sequence 27, Application US/08817926
Patent No. 6001590
GENERAL INFORMATION:
APPLICANT: Kameda, Toshinori
APPLICANT: Suda, Hisako
APPLICANT: Tamai, Yukio
APPLICANT: Imawatsu, Akihiko
APPLICANT: Kato, No. 6001590uo
APPLICANT: Sakai, Yasuyoshi
TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINI
TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,926
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/02597
FILING DATE: 12-SEP-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 234133/1995
FILING DATE: 12-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 42536/1996
FILING DATE: 29-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 081356/0112
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 4818 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Candida boidinii
US-08-817-926-27

Query Match 2.08; Score 62.6; DB 3; Length 4818;
Best Local Similarity 44.48; Pred. No. 0.001;
Matches 338; Conservative 0; Mismatches 419; Indels 4; Gaps 2;

QY 1525 ACATTAATTTGCTTCTTAATGCAATCTTAACATATTTATGTTGGAATTAAT 1584
DB 4020 AATCTTTAATGATTTCTTTTATTTGATTTTCAATTTCTGATTAATTAATTTGTTG 4079
QY 1585 AAGAGCGAATGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1644
DB 4080 TTGATTTCAATGATTTATTAATTTATTTATTTGATTTTAAAAAATTTTCTCTT -TTT 4138
QY 1645 TACATCTTTGATTTTAACTTAATCTGTTCTTCTTCCATGCTAATTAATTAATCAT 1704
DB 4139 TATTTCTTTGTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4198
QY 1705 CGAGGTAAAAAGTTTGTCTTATTTCCGATGCAATGAAGTAATTAATTAATTAATTAAT 1764
DB 4199 ATTAATTAATTTATTTACTGCAACGTTTAAAAAGATGATTAATTAATTAATTAATTAAT 4258
QY 1765 AATTTTGAATAATGTAACCTTTTACT---CATAGATTAATTAATTAATTAATTAATTAAT 1821
DB 4259 CTGAGTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4318
QY 1822 CCATTAATGACGCTCTCAACCTGATGATGCAATTTTCTGCAATTAATTAATTAATGA 1881
DB 4319 TTAATTAATTTGATTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4378
QY 1882 ATTCATGCTACTATCAATGAGAAACAGCTGATTAATTAATTAATTAATTAATTAATTAAT 1941
DB 4379 CTTTGTGTTTATCTGTTTAAATTTTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 4438
QY 1942 ATTTTGAATAATGTTAATTTCTAACAATTAATTAATTAATTAATTAATTAATTAATTAAT 2001
DB 4439 ATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4498
QY 2002 TTCCATGTTCTTAATAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2061
DB 4499 TTCTAATTTTGAACAAACCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4558
QY 2062 ATAGTGCTGAATTCAAATCTCCCAATTAATTTTGAATTAATTAATTAATTAATTAATTAAT 2121
DB 4559 TTGTTTAATTTGATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 4618
QY 2122 TTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2181
DB 4619 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4678
QY 2182 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2241
DB 4679 TGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4738
QY 2242 GTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2282


```

: REGISTRATION NUMBER: 38.241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGL1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8699
: INFORMATION FOR SEQ ID NO: 288:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 837 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG124LRP
:
: US-08-998-416-288

```

Query Match	1.9%	Score 60.6	DB 4	Length 837
Best Local Similarity	49.1%	Pred. No. 0.0017		
Matches 191, Conservative	0	Mismatches 194	Indels 4	Gaps 1

QY	1939	AAAATTTTGGAAAATGTTTAAATTTTCACAAATTTTAAATATGATGCGTAAATG	1998
Db	608	ATATATTTTATTTAATTAATTAATTAATGAAAAATATTAATTAATTAATTAATTAAT	549
QY	1999	TATTTCTATGTTCTTAAATATTTTTTTTTTATATTTAGTTATTAATACATTATGAA	2058
Db	548	AAAAATTAATTAATTAATTAATGAATATTAATTTATTAATTAATTAATTAATTAATTA	489
QY	2059	ATATATGTTGGTGAATTCAAATATCTCATTAATATTTTTTGGAAATCTACAAATTA	2118
Db	488	GAAATTAAGTTAAATTAATTAATTTTAAATATTAATCTTATAAAAGATTAATTAAT	429
QY	2119	TATTTAGCATTAACAATGCATGACAAAGTCCAAAAAAATTTGGTATGCAAGAACTCC	2178
Db	428	TCACATATATATTTTATTAATAATTAATATTTATTAATTAATAATATTAATTTACAA	369
QY	2179	AAATTTTTTTTTTTTATGGAACAAGAAATAACAGATGAAAACTATTTGTGTGGAATG	2238
Db	368	TAAATTTAACTTTATATATATTAATTAATAATATATTTTAAATAACATTAAT --- A	313
QY	2239	GAACTAGTAATATACATTAAAGCAATTTTAAAAAATTAATTAAGCTATACGCGCTCAA	2298
Db	312	ATATATTAATTAATTAATGATTAATCTATTTTATTAATTAATTAATAAATAATTAAT	253
QY	2299	GTATGTAATCTAGTAGTGTAATTAATA 2327	
Db	252	ATATTTTAATTAACATTAATTTAAATTTGAA 224	

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

1  APPLICATION NUMBER:  US/08/446,855A
2
3  FILING DATE:  06-Jul-1995
4
5  CLASSIFICATION:  435
6
7  ATTORNEY/AGENT INFORMATION:
8
9  NAME:  Milthead, Leonard C
10
11  REGISTRATION NUMBER:  29,009
12
13  REFERENCE/DOCKET NUMBER:  47-80
14
15  TELECOMMUNICATION INFORMATION:
16
17  TELEPHONE:  703-816-4000
18
19  TELEFAX:  703-816-4100
20
21  INFORMATION FOR SEQ ID NO:  1:
22
23  SEQUENCE CHARACTERISTICS:
24
25  LENGTH:  8920 base pairs
26
27  TYPE:  nucleic acid
28
29  STRANDEDNESS:  single
30
31  TOPOLOGY:  linear
32
33  MOLECULE TYPE:  genomic
34
35  US-08-446-855A-1

```

Query Match	1.9%	Score	60.4	DB	2	Length	8920
Best Local Similarity	48.8%	Pred. No.	0.0031				
Matches	163	Conservative	0	Mismatch	171	Indels	0
						Gaps	0

QY	1919	TTACATTTTATTTAAAGACAAATTTTTGGAAAATGTTATTAATTTCTAACAAATATATT	1978
Db	8751	TTTTATTTTTTTTAAATTAATTAATTTATTTAAATTTATTAATTTATTTTGGAAATTCGAA	8692
QY	1979	AAAAATGATGCCATTAATGTAATTCCTGATGTCCTAAAATATTTTTTTTTTANATTTAGT	2038
Db	8691	AAACCATTTGGTTTTATACATATACAGCAAAATTAATTAATTTGATATATATACAAAATTT	8632
QY	2039	TATATACATTTATGAAACCAATATATAGTGGTGATTCATCAATATCTCCATTAATATTTTT	2098
Db	8631	TATAAAAAATTAATCATATTAATTAATTAATGATGCATATTTATTAATAAAACATTTATTAATAT	8572
QY	2099	TGAAATCTACAATTTATTAATTTATTTAGTCATTAACAATGCTATGAAAGTTCGAAAAAAA	2158
Db	8571	TTAAACGTAAATATAAATTTAAATTAATAATTAATATATATATATATATATATATATATATA	8512
QY	2159	TTTTTGTTAACGAAGACTCCAAATTTTTTTTTTTTTTTTATGGAACAAGAAATAACAGATGAA	2218
Db	8511	TATATATATATAGTATTTATATATATTTATACGTTTCATGTTGTATNGAAGTAAGTTAAT	8452
QY	2219	AACATATTTTGTGTGGAATGGAAGTAGTAATATA	2252
Db	8451	AAATACATTAATAAGATGTTAAATCGAAATTTTA	8418

Search completed: June 11, 2003, 01:53:50
Job time : 155.509 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 01:50:19 ; Search time 459.588 Seconds
(without alignments)
9750.738 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atcttggtatattatgtg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/us07_PUBCOMB.seq:**
- 2: /cgn2_6/ptodata/1/pubpna/PCCT_NEW_PUB.seq:**
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:**
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:**
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:**
- 6: /cgn2_6/ptodata/1/pubpna/PCRTUS_PUBCOMB.seq:**
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:**
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:**
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:**
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:**
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:**
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:**
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:**
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.6	2.9	960	US-10-198-846-6381	Sequence 6381, Ap
2	82.6	2.6	15732	US-10-239-676-95	Sequence 95, Appl
3	81.6	2.5	5689	US-10-239-676-90	Sequence 90, Appl
4	80	2.5	640681	US-09-790-988-1	Sequence 1, Appl
5	79.4	2.5	960	US-10-198-846-6381	Sequence 6381, Ap
6	75.2	2.3	5979	US-10-239-676-18	Sequence 18, Appl
7	73.4	2.3	5689	US-10-239-676-90	Sequence 90, Appl
8	73	2.3	5979	US-10-239-676-18	Sequence 18, Appl
9	71.2	2.2	15732	US-10-239-676-95	Sequence 95, Appl
10	71	2.2	12405	US-10-239-676-35	Sequence 35, Appl
11	68.8	2.1	335913	US-09-754-853A-2	Sequence 2, Appl
12	68.8	2.1	335913	US-09-754-853A-3	Sequence 3, Appl
13	68	2.1	6298	US-10-239-676-64	Sequence 64, Appl
14	68	2.1	11260	US-10-239-676-20	Sequence 20, Appl
15	67.4	2.1	5917	US-10-087-464-9	Sequence 9, Appl
16	67.2	2.1	4985	US-10-094-240-10	Sequence 10, Appl
17	67.2	2.1	8588	US-10-239-676-178	Sequence 178, Appl
18	67	2.1	4985	US-10-094-240-10	Sequence 10, Appl
19	66.6	2.1	640681	US-09-790-988-1	Sequence 1, Appl

20	66.4	2.1	424	US-09-960-352-11218	Sequence 11218, A
21	66.4	2.1	7089	US-10-239-676-67	Sequence 67, Appl
22	66.4	2.1	7906	US-10-239-676-97	Sequence 97, Appl
23	65.8	2.1	3991	US-10-074-045-60	Sequence 60, Appl
24	65.8	2.1	7195	US-10-239-676-30	Sequence 30, Appl
25	65.2	2.0	11812	US-10-239-676-210	Sequence 210, Appl
26	65.2	2.0	335913	US-09-754-853A-2	Sequence 2, Appl
27	65.2	2.0	335913	US-09-754-853A-3	Sequence 3, Appl
28	64.8	2.0	6544	US-10-239-676-46	Sequence 46, Appl
29	64.8	2.0	6620	US-10-239-676-195	Sequence 195, Appl
30	64.4	2.0	473	US-09-969-376-1212	Sequence 1212, Ap
31	64.2	2.0	419	US-09-960-352-11234	Sequence 11234, A
32	64.2	2.0	17142	US-10-239-676-206	Sequence 206, Appl
33	64	2.0	9515	US-10-239-676-159	Sequence 159, Appl
34	63.8	2.0	17848	US-10-239-676-28	Sequence 28, Appl
35	63.6	2.0	11836	US-10-239-676-102	Sequence 102, Appl
36	63.2	2.0	6620	US-10-239-676-196	Sequence 196, Appl
37	63.2	2.0	11836	US-10-239-676-102	Sequence 102, Appl
38	63	2.0	11812	US-10-239-676-210	Sequence 210, Appl
39	62.6	2.0	12968	US-10-239-676-202	Sequence 202, Appl
40	62.6	2.0	15732	US-10-239-676-95	Sequence 95, Appl
41	62.4	1.9	6158	US-10-239-676-24	Sequence 24, Appl
42	62.2	1.9	6203	US-10-239-676-184	Sequence 184, Appl
43	62.2	1.9	6298	US-10-239-676-64	Sequence 64, Appl
44	62.2	1.9	11812	US-10-239-676-209	Sequence 209, Appl
45	62.2	1.9	127197	US-09-754-853A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-198-846-6381/C
Sequence 6381, Application US/10198846
Publication No. US200300997441
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinhann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: THERAPY OF BREAST CANCER
CURRENT FILING DATE: 2002-07-18
CURRENT FILING DATE: 2002-07-18
PRIORITY FILING DATE: 2001-07-18
PRIORITY FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6381
LENGTH: 960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc.feature
LOCATION: 428, 433, 450, 465, 466, 468, 471, 487, 510, 512, 513, 516,
LOCATION: 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
LOCATION: 517, 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc.feature
LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
LOCATION: 871, 874, 875, 891, 882, 895, 888, 900, 901, 902, 904, 907,
LOCATION: 917, 923, 925, 926, 927, 929, 935

OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6381

Query Match 2.9%; Score 91.6; DB 9; Length 960;
Best Local Similarity 39.5%; Pred. No. 5.7e-06;
Matches 357; Conservative 0; Mismatches 539; Indels 7; Gaps 1;

QY 1413 AAAACAAGAAAAAGAAAGATTTTCTGATTCATTCATGATCAATGATA 1472
Db ATATNAAATATAAAAAAATTTTGTNNANNAACCTAACCCTTAATNTNN 900
QY 1473 GATCTTTGGGTTACACTTCGAAGTCCCTACAGGCTGACCACTGCACTATA 1532
Db ANTTNANNAAAAAAATTTTAAANNNTAATAATTTTNNAAAACTAATNTAANNATAN 840
QY 1533 ATTCCTTTCTTTATGATCTTTTACATATTTATTTAGTGAATTTATATAGACGA 1592
Db TNATTTTNTTAAAAAANNNAAAAATTAANNNTTNTAANTTTATTAACCAAAATTTT 780
QY 1593 ACTTGATACATTCACATATTTATATATGATAGTATGATTTCCAAATACATACT 1652
Db TTTTAAAAAATTTTAAAAATTAATTAATNTAATTTTATNTAATNAAAAA 720
QY 1653 TTGGATGTTAACTTAATCTGTCTCTCTACGATATAATTAATCATCGAGTAA 1712
Db TTTNAAAAATTTTAAACANNTTTTCTTTTNNATTAATAAAAAATTTTAAATTA 660
QY 1713 AAAAAGTTTGTCTATTTTCGCGATGCATGAAGATAAACCTATGACTTTAATTTT 1772
Db NNNNTTTTNTTANANATAT-----TTAAAAAANNNNTTTTNTTTTATANTTT 607
QY 1773 GAAAATGTAACTTTTACATGATTAATACCGATGTTTGTGGCCATGATGACA 1832
Db ATAAAAAATTTTNTTTTCTNNAAAAAATTTTNTTAAANTTTTAA 547
QY 1833 GCGCTACACGCTGATGATCAATTTTCTGCAATATTAATTAAGATTCGATCTA 1892
Db AAACCTTNAATTAATAAANAATTTTNNNTTNNATTAATAAATTTTNTTAN 487
QY 1893 CTATCATATGAGAAACAGCTGATTTACATTTTAAATTAAGCAAAATTTTGA 1952
Db TTCTATTAATTAATAAANNNANNAATTAATTAATTAATAAATTTATNTTNN 427
QY 1953 ATGTTATATTTCTACATATTTATTAATATGATGCTATATGATTTCCATGCTC 2012
Db AANANAAATTTAANAANTNANNAANAATTTTANTTTAAATTAATAAATTAANTTTAAC 367
QY 2013 TTAATAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2072
Db NTCATTTNTATTTTATTAATTAATAAATTTTNTTTATATTAATAAATTTNTTTTA 307
QY 2073 ATTCATATATCTCATTAATATTTTGAATCTACAAATTAATTAATTTAGCTATA 2132
Db TTAANAANAATAAANAATNTATTAAGNNNNNTNTTAAAAAAATTAATNTANNA 247
QY 2133 CATTCATTAAGAAAGTCCAAAAAATTTGTTTACGAAGAACTTCCAAATTTT 2192
Db AATTTAATTTAATTTGNAANAATTAATTAATAAATTTTATTTATTAACNAATGTA 187
QY 2193 TATGAACAAGATATACAGATAGAAACATATTTGTTGATGAGATGATATATA 2252
Db TAAAAANNTTAANAATAGNATNTTGAANTTAATTTAATATACCTTTTAAATTA 127~
QY 2253 CATTAAGCAATTTTAAAAATTAATATACCTATACGCGCTCAAGATATGATATAGT 2312
Db ANTATTTNTANTTTAAAAAATTTAAAAAATTTTAAACCGGNNNGNCGNAGATACNTGCC 67
QY 2313 AGG 2315
Db 66 NGG 64

RESULT 2

US-10-239-676-95/c
Sequence 95, Application US/10239676
Publication No. US20030082609A1

GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239, 676
CURRENT FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1

PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01

NUMBER OF SEQ ID NOS: 228
SEQ ID NO 95

LENGTH: 15732
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-95

Query Match 2.6%; Score 82.6; DB 9; Length 15732;
Best Local Similarity 43.9%; Pred. No. 0.0005;
Matches 453; Conservative 0; Mismatches 574; Indels 6; Gaps 2;

QY 1400 AACATATAAACAAGAAAGAAAGAAAGATTTTCTGATTCATTCATGATCA 1459
Db AATATTAATAACCAAAAAAATCAAAATCAAAATCAATTAATTAATTTTCTA 6773
QY 1460 TCTAAATGATGATCTTTTGGGTACAGTTTCGAAGTCCCTCAACGCTGTACCAT 1519
Db TCTTAAATAATTAATAAACAATAATCAAAATCAAAATCAAAATCAATTAATTA 6713
QY 1520 CTGCACTATTAATTTCTTTCTTTAATGATCTTTTACATATTTATTTAGTTGCAAT 1579
Db CTTAACAAATTAATTTATTAATTAATAAACAATAATTAATAATTTTACTAAAAA 6653
QY 1580 TTAATAAGAGCACTGTAAATTAATTAATTAATTAATTAATTAATTAATTA 1639
Db TATCTTAAAAAATTAATTTTAAAAAATTAATAAATAAATAAATAAATAAATAA 6595
QY 1640 CCAATATACATCTTGGATGTTAACTTAATCTTGTCTCTACGATTAATAATA 1699
Db AAAAAACCAAAATAATTAATTAATTAATAAATAAATAAATAAATAAATAAATAA 6535
QY 1700 ATCATGAGTAAAAAAGTTTGTCTTATTTTGCATGATGATGAAGATTAACCTATG 1759
Db AATATATACCTTATTAATAATTTACATATCTTAACCTAATCACTAATCAATTA 6475
QY 1760 ACTTTAATTTTGAATAATGTAACCTTTTACTATGATTAATTAATTAATTAATTA 1819
Db TTTCAATTTAATTAATAAATAAATCAATCATATATTAATTAATAAATAAATAA 6415
QY 1820 TGCCATATGACAGCTCTACAACTGTGATAGTCAATTTTCTGCAAAATTAATAATTA 1879
Db TAAATTCATTAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6355
QY 1880 GAATTCATGCT----ACTATCATATGAAGAAAGAGCTGATATTAATTTTAA 1935
Db AAAAAAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6295
QY 1936 GACAAATTTTGAATAATGTTATTAATTTTCAACAATTAATTAATTAATTAATTAATTA 1995
Db TACATTAATTTCACTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6235

```
OY 1996 ATGATTTCTATGTTCTTAAATATTTTTTTTATTTTACTTATTAATACATTATGAA 2055
    || || || || || || || || || || || || || || || || || || || || ||
Db 6234 ATTATATTATAAATTAATATCTATTATTACTTATTAATAATACATTAAATTTTTTCTT 6175
OY 2056 CCAATTAATAGTTGGGAATTCAAATATCTCCATTAAATTTTGGAAATCTCAAAATAT 2115
    || || || || || || || || || || || || || || || || || || || || ||
Db 6174 CTCTCTAAACACACATTAACATATATCTTAAATATCATATTTCTATTAATAAATATAT 6115
OY 2116 TAATATTAGTCAATTAACATATGATAGTCAAGTCCAAAAAATTTGTTACAGAACT 2175
    || || || || || || || || || || || || || || || || || || || || ||
Db 6114 AATTAATAATCAAAAAACATACATATAAATAATTAACAAATTCACATTAAATTAATAAAT 6055
OY 2176 TCCAAATTTTTTTTTTTATGAGACAGAAATTAACAGATAGAAATCTATTTGTTGTA 2235
    || || || || || || || || || || || || || || || || || || || || ||
Db 6054 TCACACATCTCTCAATTACTTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5995
OY 2236 ATGAGATAGTAAATATACATTAGCAAAATTTTAAAAATTAATATAGCCTATACGGCTC 2295
    || || || || || || || || || || || || || || || || || || || || ||
Db 5994 ATTTTCAAAACATTAATTAACAAACTTACTTAAAAAATTTATCTTAAATAAATAATTT 5935
OY 2296 AAGATATGTTATCTAGTATGTTATTAATATATGATGCTGCAATTCAGAACTTGACAA 2355
    || || || || || || || || || || || || || || || || || || || || ||
Db 5934 AAAATCCACAGTTTAAACCTCTCTCAATTAATCTATAAATAATTAATAAATAAATAA 5875
OY 2356 CAATGAAAAAGCAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2415
    || || || || || || || || || || || || || || || || || || || || ||
Db 5874 TATTAACAAAACATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 5815
OY 2416 TTCTGATCTATTGA 2428
    || || || || || || || || || || || || || || || || || || || || ||
Db 5814 CCTTATCTTTTCA 5802
    || || || || || || || || || || || || || || || || || || || || ||
```

RESULT 3

```
US-10-239-676-90/c
; Sequence 90, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 90
; LENGTH: 5689
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-239-676-90
```

```
Query Match
Best Local Similarity 49.1%; Score 81.6; DB 9; Length 5689;
Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;
OY 1824 ATAAATGACAGCTCTACAACTGATGATGCAATTTTCTCGCAAAATTAATTAAGCAAT 1883
```

```
Db 2110 AAAATCGCAATCCGCAACCGCAATTAACAAAAAATTTTCCCGCAAAAAAATAAAAA 2051
    || || || || || || || || || || || || || || || || || || || || ||
OY 1884 TCAATGCTACTCAATAGAAAGACAGCTGATTAATCAATTAATTAAGACAAAT 1943
    || || || || || || || || || || || || || || || || || || || || ||
Db 2050 TATTAATAAATAAACAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1991
    || || || || || || || || || || || || || || || || || || || || ||
OY 1944 TTTGAAAAAATGTAAATTTCTAACATATTAATTAATAATATAGAGCCATATATAT 2003
    || || || || || || || || || || || || || || || || || || || || ||
Db 1990 ATTATTTAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1931
    || || || || || || || || || || || || || || || || || || || || ||
OY 2004 CCTATGCTTAAATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 2063
    || || || || || || || || || || || || || || || || || || || || ||
Db 1930 NATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1871
    || || || || || || || || || || || || || || || || || || || || ||
OY 2064 AGTTGCT -GAATCAAAATCTCCCAATTAATTTTGGAAATCAAAATTTATATAT 2122
    || || || || || || || || || || || || || || || || || || || || ||
Db 1870 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1811
    || || || || || || || || || || || || || || || || || || || || ||
OY 2123 TAGCAATTAACATGCAAT -AGAAAGTCCAAAAAATAATTTGTTAACAGAACTCCAAA 2181
    || || || || || || || || || || || || || || || || || || || || ||
Db 1810 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1751
    || || || || || || || || || || || || || || || || || || || || ||
OY 2182 TTTTTTTTTTTATGGAACAAGAAATTAACAGATAGAAACTATTTGTTGGAATGAA 2241
    || || || || || || || || || || || || || || || || || || || || ||
Db 1750 TAAATTAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1691
    || || || || || || || || || || || || || || || || || || || || ||
OY 2242 GTAGTAATTAATTAAGCAATTTTAAATAATTAATTAATTAATTAATTAATTAATTA 2301
    || || || || || || || || || || || || || || || || || || || || ||
Db 1690 TAAACAATTAATAAATAAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTA 1631
    || || || || || || || || || || || || || || || || || || || || ||
OY 2302 TGTATCTAGTAGTGAATTAATA -ATGCATGCTGCAATTAATTAATTAATTAATTA 2360
    || || || || || || || || || || || || || || || || || || || || ||
Db 1630 TATTAATCTATCAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1571
    || || || || || || || || || || || || || || || || || || || || ||
OY 2361 AAAACGAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
    || || || || || || || || || || || || || || || || || || || || ||
Db 1570 ATAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1531
    || || || || || || || || || || || || || || || || || || || || ||
```

RESULT 4

```
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1
```

```
Query Match
Best Local Similarity 45.8%; Score 80; DB 10; Length 640681;
Matches 395; Conservative 0; Mismatches 455; Indels 12; Gaps 3;
```

```
OY 1559 ATATTATTGTTAGTTGAATTAATAAGCAAGCACTGTAACTATTAATATTATTATT 1618
    || || || || || || || || || || || || || || || || || || || || ||
Db 324686 ATATTGTTAGTATCTTTTATTAATTAACCAATTAATTAATTAATTAATTAATTAAT 324745
    || || || || || || || || || || || || || || || || || || || || ||
OY 1619 AGATTAAGTATGATGATTAATTCCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1678
    || || || || || || || || || || || || || || || || || || || || ||
```

```

Db 324746 AATCTTTTACAAAGAAATTAATATTTTCTTAACTAATTCATTTTGATC 324805
QY 1679 CTTCCACGGTAATTAATATTAATCAGGTAAGGTAAGGTTTGTCTATTTCCGAT 1738
Db 324806 TCTTCATCATGTAATAATCAATCTTCTTCTAATATCAATATCTTTAATATAT 324865
QY 1739 GCATGAAGGTAACCAATGACTTATTTTGAATGTAACCTTTTACTCATAGA 1798
Db 324866 TGTGGGGTATTTTATGTCATTAATAAACTTAATTTTAAAGATTTTCTCT 324925
QY 1799 TTAATACGATATTTTGTGCGATAGACAGCCTACAGCCTGATAGCATTT 1858
Db 324926 GTCAAGAAATATATATTTTCTCGAATATGAATATTAATAGCAATACGTATATTT 324985
QY 1859 TTTCTCGAATATTAATTAATAGATTCATGCTACTATCAATAGAGAAACAGCTGATA 1918
Db 324986 TTTTGTGATTAATTAATTAATTAATAGATCTAGTATTTTGTATTTAAATTAATA 325045
QY 1919 TTAATTTTAATTAATTAAGCAAAATTTTGAAGAAATGTAATTTTCAACATATATT 1978
Db 325046 ATTTTGTGTTTAAATTTTAAATTTTGTATATTTTCAATAGACGATTTCTTTAAAT 325105
QY 1979 AATATATGATGCTAATATGATTTCTCTA----TGTCTAAATATTTTCTTAAAT 2033
Db 325106 AAGTATATATTTTATTAATTAATTAATTAATTAATTTTATTCATATTAATATTTTAA 325165
QY 2034 TTAGTATATAATCATTAATGACCAATTAATAGTGTGATTCATCAATATCTCATTAATA 2093
Db 325166 ATTTAAATATATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325225
QY 2094 TTTTGTGAATCTACAAAT-----TATTAATTTTACTCAATTAACATGATGAAAGTT 2148
Db 325226 ATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325285
QY 2149 CCAAAATTTTGTGTAACAGAACTTCCAAATTTTGTGTAAGCAAGAAATA 2208
Db 325286 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325345
QY 2209 ACAGATTAAGAACTATTTTGTGTGGAAGGAAAGTAAATTAATTAATTAATTAATA 2268
Db 325346 ACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325403
QY 2269 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2328
Db 325404 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325463
QY 2329 GCATGTCGATTCAGAAATGGAACAACATGAAGAAAGCAATTAATTAATTAATA 2388
Db 325464 GATTTGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325523
QY 2389 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2410
Db 325524 TTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 325545

```

```

RESULT 5
US-10-198-846-6381
; Sequence 6381, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steilmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6381
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
; LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
; LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
; LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427.
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 428, 433, 450, 465, 468, 471, 487, 510, 512, 513, 516,
; LOCATION: 517, 518, 528, 530, 534, 539, 553, 559, 575, 580, 588,
; LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
; LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc-feature
; LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
; LOCATION: 871, 874, 875, 891, 892, 895, 898, 900, 901, 902, 904, 907,
; LOCATION: 917, 923, 925, 926, 927, 929, 935
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6381

```

Query Match 2.58; Score 79.4; DB 9; Length 960;

Best Local Similarity 40.9%; Pred. No. 0.00057;

Matches 312; Conservative 0; Mismatches 445; Indels 5; Gaps 2;

```

QY 1532 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1591
Db 70 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1591
QY 1592 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1649
Db 130 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1649
QY 1650 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1709
Db 190 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1709
QY 1710 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1769
Db 250 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1769
QY 1770 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1826
Db 310 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1826
QY 1827 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1886
Db 370 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1886
QY 1887 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1946
Db 430 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 1946
QY 1947 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2006
Db 490 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2006
QY 2007 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2066
Db 550 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2066
QY 2067 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2126
Db 610 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2126
QY 2127 AATGCTTCTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 2186

```


Db 670 AAAAANTTTTATNNNAANAAAAAANNTGNTAAATTTTAAATTTTTTTT 729

QY 2187 TTTTATGACAGAAATACAGATAGAAACATTTGTTGGATGAGACATG 2246

Db 730 TNNATAAAAAATTAANTATTAATTTTAAAAAATTTTTTAAAAAATTTTGG 789

QY 2247 AATATACATTACCAATTTTAAAAATATATATACCCATA 2288

Db 790 TTTAATTAANTTANAAAAANNNTTAAATTTNNNTTTTTTANA 831

RESULT 6

US-10-239-676-18

Sequence 18, Application US/10239676

Publication No. US20030082609A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003

CURRENT APPLICATION NUMBER: US/10/239,676

PRIOR FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8

DE 10019173.8

DE 10032529.7

DE 10043826.1

PRIOR FILING DATE: 2001-04-06

2000-04-06

2000-04-07

2000-06-30

2000-09-01

NUMBER OF SEQ ID NOS: 228

SEQ ID NO 18

LENGTH: 5979

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-18

Query Match

Best Local Similarity 44.7%; Score 75.2; DB 9; Length 5979;

Matches 388; Conservative 0; Mismatches 468; Indels 12; Gaps 2;

QY 1554 TTACATATTTATGTAGTGTAGTAAATTAATTAAGACGCACTGTAACATTCATATTT 1613

Db 1459 TAAAGTAAATGTATATATTTTATTTAGAAATTTATTTTGTGTTTTTATTA 1518

QY 1614 AATATGATCTAGTATGTATTTCCAAATACATCTTGATGTTAACTTAATCT 1673

Db 1519 AATTTTAAATTTATATATATTTTGAATAATATATATTTTAAAAAAGAAATTTT 1578

QY 1674 TGTTCCTGCTAGCTATTAATTAATTAATCAATGAGCTAAAAAGTTTGTCTTATTTTC 1733

Db 1579 TTATATATTAATATATATTTTATTTATGATTTAGCTAAATTTTAAAGATGCTT 1638

QY 1734 GCGATGATGAGAGATAACCTAATGACTTTAATTTTGAATAATGTAACCTTTACTC 1793

Db 1639 TAAATGACATATTTTAAATTAAGATTTTAAAGAAATTTTACGTTAATTAAGAA 1698

QY 1794 ATAGTTAATTAACCTATGTTTGTGTCATATGACAGCCTCTACACGTGTATGTC 1853

Db 1699 ATATATATTTTAAATTTTATTTTAAATTTTAAATTAAGTATATTTATTTTATA 1758

QY 1854 AATTTTCTGCAATATTAATTAATTAAGATTAATCAATGCAATCAATGAGAAACAGCT 1913

Db 1759 ATGTATTAATTAATTAATTAATTAAGATTTTAAAGTATGTTTAAAGTAA 1818

QY 1914 GAGTATTAATTAATTTTAAAGCAAAATTTTGAATAATGTTATATTTTAAACAATA 1973

Db 1819 TTTTAAATTAATTAATTAATTAAGATTTATGAAATTAATTAATTTTAAAGAAATAT 1878

QY 1974 TTATTA - AATATGATGCCATATATGATTTTCCATGCTCTTAAATTTTTTTTAT 2031

Db 1879 ATTTTAAATTAATTTTATTTATTTAGATGATGATATTTATATTTTAAAAAATTTAAT 1938

QY 2032 ATTTAGTTAAATCATATATGACCAATTAATGTTGGATTCAAATTCCTCAATTA 2091

Db 1939 TTTTATTTATTTATTAATTTTGTGAAATTAATGTTT-----TTTTTATGGAAG 1988

QY 2092 TATTTTGAATCTACAAATTTTAAATTTAGTCAATACATGATAGAAAGTGTCA 2151

Db 1989 AATTTATTTTATGATTTATTAATTAATTAATTAATTTTATTTAGTAAATTTAT 2048

QY 2152 AAAAAAATTTTCTTAACAGAACTCCAAATTTTTTTTTTATGCAACAGCAATAACA 2211

Db 2049 AATTTAAATGTTTAAATTTTATTTATTTATATTAAGATTAATTTCTGATGACAAATA 2108

QY 2212 GATGAAACATATTTGTTGTCGATGAGATGAGATGATATATACATTAACCAATTTTAAA 2271

Db 2109 TATTTTAAATTTTATTTTGAATTTAATTAAGATTTAATTTTATTTATGTTAAATTA 2168

QY 2272 AATTAATTAAGCTATACGCGCTCAAGATGTTATCTAGTAGTATTAATTAATGCA 2331

Db 2169 AATGCTGTCGTATATTAATTAATGATGATGTTGTAGATTTGATATGATATATAA 2228

QY 2332 TGTGCGATTCAGATTTGGACACATAGAAACGCAATTAATTAATTAATTAATA 2391

Db 2229 TAGTATGATTAATTAATTAATTTGAATTTGATATTTTATTTTATGTAATAATGCG 2288

QY 2392 AATAAAAATTTGAGTAAATGTTTCT 2419

Db 2289 AAAATGAAATTAAGAAATTTTGTCTT 2316

RESULT 7

US-10-239-676-90

Sequence 90, Application US/10239676

Publication No. US20030082609A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003

CURRENT APPLICATION NUMBER: US/10/239,676

PRIOR FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8

DE 10019173.8

DE 10032529.7

DE 10043826.1

PRIOR FILING DATE: 2001-04-06

2000-04-06

2000-04-07

2000-06-30

2000-09-01

NUMBER OF SEQ ID NOS: 228

SEQ ID NO 90

LENGTH: 5689

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

NAME/KEY: unsure

LOCATION: (1818, 1930, 1931, 1934, 1947, 1948, 1966)

US-10-239-676-90

Query Match

Best Local Similarity 44.8%; Score 73.4; DB 9; Length 5689;

Matches 304; Conservative 0; Mismatches 374; Indels 1; Gaps 1;

QY 1518 ATCTGACATTAATTAATGCTTCTTAAATGACATCTTTAAACATATTTATTTAGTTGA 1577


```

Db 5453 ATGGAAATTAATGATAGAAATTTTAAATGAAATGAAATGAAAGAAAGAAAGAAATG 5511
QY 1591 GAACCTTGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1650
Db 5512 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5571
QY 1651 CTCTGATGTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1709
Db 5572 GAATATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5631
QY 1710 -----TAAAAAAGTTTGTCTATTTTCCGATGCAATGAAGATTAACCTAATGAC 1761
Db 5632 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5691
QY 1762 TTTAATTTTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1821
Db 5692 ATAGTTATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5751
QY 1822 CCATTAATGACAGCTCTACAACTGATGATGATGATGATGATGATGATGATGATGATG 1881
Db 5752 AGTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5810
QY 1882 ATTCATGCTACTATCAATGAAGAAACACCTGATGATGATGATGATGATGATGATGATG 1941
Db 5811 TTTAAATTTTAAAGTTTAAATTTTAAATTTGAAAGGAGGAAATTTTAAATTTATGT 5870
QY 1942 ATTTTGAATAATGTTAATTTCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2001
Db 5871 ATGATTTAGATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5930
QY 2002 TTCCATGTTCTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2061
Db 5931 AAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 5990
QY 2062 ATAGTTGCTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2121
Db 5991 TTTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 6050
QY 2122 TTATGCTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2181
Db 6051 TAAATGATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 6110
QY 2182 TTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2239
Db 6111 TTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 6170
QY 2240 AATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2281
Db 6171 ATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 6212

```

```

RESULT 10
US-10-239-676-35
; Sequence 35, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 3013.1003
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US/10/239, 676
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30

```

```

; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 35
; LENGTH: 12405
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (7895)
; US-10-239-676-35

```

```

Query Match 2.2%; Score 71; DB 9; Length 12405;
Best Local Similarity 45.0%; Pred. No. 0.036;
Matches 347; Conservative 0; Mismatches 420; Indels 4; Gaps 2;

```

```

QY 1512 GTAACATGTCGCACTTAATTAATTTGCTTTAATGCACTTTAATTAATTAATTTA 1571
Db 190 GGAGCCGTTTGGCTTTTAAAGTAACTAATTTAATTAATTTAATTTAATTTAATTTA 249
QY 1572 GTTGAATTAATTAAGCCGACCTGTAATTAATTAATTAATTAATTAATTAATTAAT 1631
Db 250 TTGCTAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 309
QY 1632 TGAATTTCCAA--ATACATACCTTGGATGTTTAACTAATCTTCTTCTTCTTCTT 1689
Db 310 TGTGTTGTTTATTTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 369
QY 1690 ATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 1749
Db 370 TTAGGTATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 429
QY 1750 AAACCTAATGCTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1809
Db 430 TTTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 489
QY 1810 ATGTTTGTGTCATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1869
Db 490 AAGATTTTGTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 549
QY 1870 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 1927
Db 550 TTGTTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 609
QY 1928 AATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1987
Db 610 TAAATTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 669
QY 1988 TGCCCTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2047
Db 670 TGTATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 729
QY 2048 ATTAATGAACAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2107
Db 730 ATTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 789
QY 2108 CAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2167
Db 790 GTTTTGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 849
QY 2168 CAGAACTTCCAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2227
Db 850 ATTTGATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 909
QY 2228 GTTGTGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2278
Db 910 GTGTTTAAATTTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 960

```

```

RESULT 11
US-09-754-853A-2/c
; Sequence 2, Application US/09754853A

```


Db 99159 A---AAATATTTAACTTTTAAATTTAGTTTAAATTAAGTATTTAAAAA 99103
Qy 1970 AAATATTTAAATATGATGCTATATGATTTCTGATTTCTAAATATTTT 2029
Db 99102 AAAACATATTAATATCATATGATTTATTTTATTTTATTTTACAAACAACTT 99043
Qy 2030 ATATTTAGTTTAAATATCATATGATTTATTTTATTTTATTTTATTTTATTT 2089
Db 99042 CAATATCATTTTAAATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 98983
Qy 2090 AAATATTTTGAATATCAAT---TATTAATATTTAGCATTAATATGATTAAG 2146
Db 98982 AAATATCTAGTAAATATTAATTTTATTTTAAATTAATTAATTAATTTTAAATTT 98923
Qy 2147 TTCCAAAAAATTTTGTATACAGAACTTCCAAATTTTATTTTATGGAACAGAAA 2206
Db 98922 TTTTAAAGAAATGAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTA 98863
Qy 2207 TAACAGATGAAATCTATTTTGTGGAATGGAAGTATTAATTAATTAATTAATTT 2266
Db 98862 AAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 98803
Qy 2267 TAAATATTTATAGCTATAGCGCTCAAGATGATTTATCTAGTATGATTAATA 2326
Db 98802 ACAGCTTTAAATATCATATGATTTAAATTAATTAATTAATTAATTAATTAATTT 98743
Qy 2327 ATGATGCTGATGATTCAGAAATGGAACATGAAACGAAATTAATTAATTTACTTTA 2386
Db 98742 ATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98683
Qy 2387 AAATTAATA 2396
Db 98682 AAATTAATA 98673

RESULT 13
US-10-239-676-64/c
; Sequence 64, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 64
; LENGTH: 6298
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-64

Query Match 2.1%; Score 68; DB 9; Length 6298;
Best Local Similarity 48.8%; Pred. No. 0.087;
Matches 273; Conservative 0; Mismatches 280; Indels 7; Gaps 3;

Qy 1850 AGTCATTTTTCGAAATTAATTAATTTGCAATTCATCTCATATAGAAAC 1909
Db 4117 AATACATTTTAAATTAATTAATTAATTAATTAATTTATCTATTAATTAATTT 4058

Qy 1910 AGTCATATTCATTTTAAATTTAAAGCAAAATTTTGAATAATTTATTAATTTTAC 1969
Db 4057 TCCTTTTAAATATCATATTTATTAATA---AAATATTTAAATTAATTAATTAATTTAT 4001
Qy 1970 AAATATTTAAATATGATGCTATTAATGATTTCTGATTTCTTAAATATTTT 2029
Db 4000 ACCCTTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 3941
Qy 2030 ATATTTAGTTTAAATATCATATGATTTATTTTAAATTAATTAATTAATTTATTT 2089
Db 3940 ATATTTATTAATCATATTAATTAATTAATTAATTAATTAATTAATTAATTTATTT 3881
Qy 2090 AAATATTTTGAATATCA---AATTAATTAATTTTATGATTAATTAATTAATTT 2148
Db 3880 TATATTTATTAATATCATATTAATTAATTAATTAATTAATTAATTAATTTTAAATTT 3821
Qy 2149 CCAAAAAAATTTTGTATACAGAACTTCCAAATTTTATTTTATGGAACAGAAATA 2208
Db 3820 TAACTTAAAAATATTAATTAATTAATTAATTAATTAATTTTATTAATTAATTTATCT 3762
Qy 2209 ACAGATGAATATCTATTTGTGGAATGGAAGTATTAATTAATTAATTAATTTA 2268
Db 3761 --CTACTTTAACTTTTAAATTAATCCAAAAATTTTAAATTAATTTTATCTATTA 3704
Qy 2269 AAAAATTAATTAAGCTATACGCTCAAGATGATTTATCTAGTATGATTAATTAAT 2328
Db 3703 CAAATATCTATCTTTAAATATGCAAAATTAATTAATTAATTAATTAATTTTAAATTTA 3644
Qy 2329 GCATGCTGATTCAGAAATGGAACATGAAACGAAATTAATTAATTTACTTTAA 2388
Db 3643 TTACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCTATCACTATCTATTAAT 3584
Qy 2389 ATAAATTAATTTGACTAA 2408
Db 3583 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAAAT 3564

RESULT 14
US-10-239-676-20/c
; Sequence 20, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 20
; LENGTH: 11260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-20

Query Match 2.1%; Score 68; DB 9; Length 11260;
Best Local Similarity 46.4%; Pred. No. 0.11;
Matches 256; Conservative 0; Mismatches 295; Indels 1; Gaps 1;

QY 1839 ACACAGCTGGTATTCATTTTCTTCGAAATATTAAATTAGAAATTCGAATGCTATATCA 1898
 Db 6329 AAATATTTCACAAACCTATTTTTTAACAATATATAAACCACTTTATTTTTCACAAATTCCTACCA 6270
 QY 1899 ATAGAGAAGAACAGCTGAGTATTTACATTTTATTTAAGACAAATTTTGTAAAAATGTTA 1958
 Db 6269 TAAAAAATAAATAATATAATATATTCACCTATATTTAAATTCACATATTAATTTCTA 6210
 QY 1959 TAATTTCTACAAATATTATTAATAAATATGATGCTATATATTTTCCATGTTCTTAAAA 2018
 Db 6209 CCAAAATATATATACAAATATTTTATATTTTAAACCACTTTTTCACATTTTAAATATATAAAA 6150
 QY 2019 TATTTTTTTTTATTTAGTTATTAATAATACATTAATGAAACCAATATAGTTGGTAATTCAA 2078
 Db 6149 TCTTTTATACATATTCTATTATATATATATTTATAT -ATTTTTTCCATACAAAACAA 6091
 QY 2079 ATATCTCCATTAATATTTTTTTGCAATCTCAAAATTTTATATTTTATTTAGTCATATACAAATG 2138
 Db 6090 ATCATCTACTTATATATATATATTTAACTAAATATACCCATTTTATTTTAAATATCAATTT 6031
 QY 2139 ATAGAAAGTCCAAAAAAATTTTGTATAACAGAACTCCAAATTTTTTTTTTATATGGA 2198
 Db 6030 AAAAAACATTAATACACTTAACATTAACAAACATTAACATATTTTAAATTTTAAATTC 5971
 QY 2199 ACAGAAATATACAGATAGAAAACCTATTTTGTGTGGAATGGAAGTATATATACATTA 2258
 Db 5970 CTCTAAATTAATTTCTATCAATCCCATTAATATTTCTATTAATAAATAATTTCAATATACA 5911
 QY 2259 GCAAAATTTTAAAAAATATATATAGCCTATATACGCGCTCAAGATGTATCTATGATAGTGT 2318
 Db 5910 ATAACTTTAAACAATAATACCATATCTTAATATCTCCCTAATCATTTTACATTAACATAT -5851
 QY 2319 AATTATATATGATGATGGATTCAGAAATTTGGACACATGAAACGAATTTAAATAT 2378
 Db 5850 TAAAAACATTAATAAATTCCTACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 5791
 QY 2379 TAACTTTAAAT 2390
 Db 5790 ATATTTTAAACAT 5779

RESULT 15
 US-10-087-464-9/c
 : Sequence 9, Application US/10087464
 : Publication No. US20030059436a1
 : GENERAL INFORMATION:
 : APPLICANT: Chisholm, Athar
 : APPLICANT: Oh, Steven
 : APPLICANT: Liu, David
 : APPLICANT: Goel, Vikas
 : APPLICANT: Li, Xuerong
 : TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
 : CURRENT APPLICATION NUMBER: US/10/087,464
 : PRIOR FILING DATE: 2002-03-01
 : PRIOR APPLICATION NUMBER: US 06/272,930
 : NUMBER OF SEQ ID NOS: 59
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 9
 : LENGTH: 5917
 : TYPE: DNA
 : ORGANISM: Plasmodium falciparum
 : US-10-087-464-9

Query Match	2.1%;	Score 67.4;	DB 9;	Length 5917;
Best Local Similarity	48.7%;	Pred. No. 0.11;		
Matches 211;	Conservative 0;	Mismatches 221;	Indels 1;	Gaps 1

D

QY 1848 5867	<pre>ATAGCAATTCTTCGCAATAATTAATAGCATCGTACTAATCATAAGAGA AAAAAAAATTATGTATTATTAAGTTATTTATATCTGCAGTGTTTAAAA</pre>	1907 5808
------------------------	--	------------------

QY	1908	ACAGCTGGCTTTACATTTTAAATTTAAAGCAAAATTTTGAAAAATGTTATATTTCTA	1967
Db	5807	AAAAAAAAATGCTTTTATATGTTACGCTAAAAAAAATAATATATATATATATATGTA	5748
QY	1968	ACAAATATTATTAAATATGATGCGCTATATATGCTATTTCCATGCTCTTAAATATTTTTT	2027
Db	5747	AATATATATATACATATATAAATAAATAATATATATATATATATATATATATATATAT	5688
QY	2028	TTATATTTAGTTATTAATACATTTATGAACCAATATATAGTTGGTGAAATTCAAATATCTCA	2087
Db	5687	ATATATATAAATTTACGTACTTTTTTTTATTTTAAAAAGTTTWTAT -AATCGATATA	5629
QY	2088	TTATATTTTTTGAATCTACAATTTATATATTATTTAGTCATATAACAATGCATGAAGT	2147
Db	5628	TAAATATTTAAATATGCACTTGAAGAAAGAAAGAAAAAATAATATATATATATATCTGA	5569
QY	2148	TCCAAAAAAATTTTGTTAACAGAACTTCCAAATTTTTTTTTTTTATGAAACAGAAAT	2207
Db	5568	TAAATGTATTAATTTATAGATATATATTTTTTTTTTTTTTTTTTTTATATATATAATTT	5509
QY	2208	AACGATAGAAACTATTTTGTGTTGGATGGAAGTAGTAATATATACATTAAAGCAATTTT	2267
Db	5508	ATATTTTTTAAATATATTTTGACATGACTTAATAATACATATTCCTGTAATTACTAAAT	5449
QY	2268	AAAAAATATATATA	2280
Db	5448	ATATATATATATATA	5436

Search completed: June 11, 2003, 08:07:41
Job time : 463.588 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 7692.77 Seconds

(Without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgttggtatattatgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*

```

1: /cgn2_6/ptodata/1/pna/US0600.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097C.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099A.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B.COMB.seq:*

```

Result No.	Score	Query Match	Length	DB ID	Description
1	3202	100.0	6888	1	PCT-US500-03820-1
2	3202	100.0	6888	19	US-09-502-426-1
3	3134.2	97.9	84196	27	US-09-692-412-94
4	3134.2	97.9	84196	31	US-09-803-736-1074
5	960.4	30.0	1699	18	US-09-451-320-2050
6	125	3.9	826	1	PCT-US99-228538-3486
7	118	3.7	84196	27	US-09-692-412-94
8	118	3.7	84196	31	US-09-803-736-1074
9	105.6	3.3	1016	64	US-60-207-458-33653
10	102.4	3.2	1326	26	US-09-663-779-8205
11	98.2	3.1	1326	26	US-09-663-779-8205
12	97.6	3.0	810	26	US-09-666-355A-6400
13	97.6	3.0	22350	68	US-60-245-228-17
14	94.6	3.0	742	26	US-09-663-779-3024
15	91.6	2.9	960	41	US-10-158-846-6381
16	91.6	2.9	464387	35	US-09-948-941-643
17	89.2	2.8	98573	68	US-60-248-505-4421
18	88.4	2.8	1368	38	US-10-015-127-2421
19	87.6	2.7	908	56	US-60-126-265-2038
20	87.6	2.7	27978	64	US-60-207-423-22
21	87.6	2.7	32768	64	US-60-207-423-23

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

22 87.6 2.7 32768 64 US-60-207-905-22 Sequence 22, Appl
23 87.6 2.7 32768 64 US-60-207-905-23 Sequence 23, Appl
24 87.6 2.7 32768 64 US-60-207-905-24 Sequence 24, Appl
25 87.6 2.7 32768 64 US-60-208-837-26 Sequence 26, Appl
26 87.6 2.7 32768 64 US-60-208-837-27 Sequence 27, Appl
27 87.6 2.7 32768 64 US-60-208-837-28 Sequence 28, Appl
28 87.4 2.7 32768 64 US-60-207-458-34124 Sequence 34, Appl
29 87.4 2.7 28054 65 US-60-212-657-88 Sequence 88, Appl
30 87.4 2.7 32768 65 US-60-213-181-33 Sequence 33, Appl
31 87.4 2.7 32768 65 US-60-213-181-34 Sequence 34, Appl
32 87.4 2.7 32768 65 US-60-213-181-35 Sequence 35, Appl
33 87.4 2.7 158001 42 US-10-211-179-11 GENERAL INFO:MART
34 87 2.7 868 25 US-09-654-617-367906 Sequence 367906,
35 87 2.7 868 27 US-09-684-016-367906 Sequence 367906,
36 87 2.7 1368 38 US-10-015-127-2421 Sequence 2421, Ap
37 87 2.7 192014 68 US-60-248-823-33 Sequence 33, Appl
38 87 2.7 755 24 US-09-634-3068-128109 Sequence 128109,
39 86.8 2.7 755 38 US-10-027-632-128109 Sequence 128109,
40 86.8 2.7 755 38 US-09-404-520-128109 Sequence 128109,
41 86.6 2.7 720 18 US-09-404-520-128109 Sequence 128109,
42 86.6 2.7 961 56 US-60-126-265-3039 Sequence 3039, A
43 86.4 2.7 32768 65 US-60-213-177-227 Sequence 227, App
44 86.4 2.7 38678 35 US-09-948-947-136 Sequence 136, App
45 86.4 2.7 38684 35 US-09-948-947-154 Sequence 154, App

ALIGNMENTS

RESULT 1
PCT-US00-03820-1
Sequence 1, Application PC/TUS0003820
GENERAL INFORMATION:
APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA.
TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 2225-0001 40
CURRENT APPLICATION NUMBER: PCT/US00/03820
CURRENT FILING DATE: 2000-02-11
EARLIER APPLICATION NUMBER: 60/119,657
EARLIER FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: 60/119,658
EARLIER FILING DATE: 1999-02-11
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 6888
TYPE: DNA
ORGANISM: Arabidopsis sp.
PCT-US00-03820-1

Query Match 100.0%; Score 3202; DB 1; Length 6888;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGATATATATGTTGGTTGGTTGACGTACATATAATTTCTGTTCTGCT 60
DB 1 ATGTGGATATATATGTTGGTTGGTTGACGTACATATAATTTCTGTTCTGCT 60
QY 61 TATTCGTTCACATGATTTGAGTTGGTTGCTCAATTTGATTCACAGATATAATTT 120
DB 61 TATTCGTTCACATGATTTGAGTTGGTTGCTCAATTTGATTCACAGATATAATTT 120
QY 121 AAAATTCATTTAAATATTTTACAGTAATTAATTAATTTTACATTTGATTTATACAA 180
DB 121 AAAATTCATTTAAATATTTTACAGTAATTAATTAATTTTACATTTGATTTATACAA 180
QY 181 AATATCTATCTTGGATATATGAGAAATATGAGTTTGAATTTATATATATAGGAAA 240
DB 181 AATATCTATCTTGGATATATGAGAAATATGAGTTTGAATTTATATATATAGGAAA 240
QY 241 TAATTCATTCATTTGTTGGATTTACACAGTTAAGTTTGTCTTTCTTTGTTATATGT 300
DB 241 TAATTCATTCATTTGTTGGATTTACACAGTTAAGTTTGTCTTTCTTTGTTATATGT 300

QY 301 ATATGACTAAATCAAAAAAGATATGATTAAGATTAATTTCTGTTATGACCCCA 360
DB 301 ATATGACTAAATCAAAAAAGATATGATTAAGATTAATTTCTGTTATGACCCCA 360
QY 361 AAAAAAAAAAAAAACCAAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 420
DB 361 AAAAAAAAAAAAAACCAAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 420
QY 421 AGGTTATTTGATCATATATATGATGATGATTTCTGTTATGATTAAGATTTCTTAC 480
DB 421 AGGTTATTTGATCATATATATGATGATGATTTCTGTTATGATTAAGATTTCTTAC 480
QY 481 CAATTTAAATTTGCAATTCATATCTCTTGAATTAATTAATTAATTAATTAATTAAT 540
DB 481 CAATTTAAATTTGCAATTCATATCTCTTGAATTAATTAATTAATTAATTAATTAAT 540
QY 541 GTTTATTCGATCAGCCCAATGATGATTTGATTTGATTTGATTTGATTTGATTTGAT 600
DB 541 GTTTATTCGATCAGCCCAATGATGATTTGATTTGATTTGATTTGATTTGATTTGAT 600
QY 601 AGAGTATTTGAGAAAAACCAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 660
DB 601 AGAGTATTTGAGAAAAACCAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 660
QY 661 TGAATTAATTTCAAGAGAAATTAAGATGACAAACCAAAAGTTGTGAATTAATGCTCT 720
DB 661 TGAATTAATTTCAAGAGAAATTAAGATGACAAACCAAAAGTTGTGAATTAATGCTCT 720
QY 721 GCCAGCTTCTCCACATATATGACACCTATTGGATTTCTGATTTCTGATTTCTGAT 780
DB 721 GCCAGCTTCTCCACATATATGACACCTATTGGATTTCTGATTTCTGATTTCTGAT 780
QY 781 TTGCGATTAAGATTTGAAAAATTTTATTTGATGATGATGATGATGATGATGATG 840
DB 781 TTGCGATTAAGATTTGAAAAATTTTATTTGATGATGATGATGATGATGATGATG 840
QY 841 GGTATTTGCAATATCTCTGTTTAAACATATTTGTCTTCTTCTTCTTCTTCTTCT 900
DB 841 GGTATTTGCAATATCTCTGTTTAAACATATTTGTCTTCTTCTTCTTCTTCTTCT 900
QY 901 TAATCTATATTTATGCGGATATATGATTAACATATATATATATATATATATAT 960
DB 901 TAATCTATATTTATGCGGATATATGATTAACATATATATATATATATATATAT 960
QY 961 ACCATTTGAAATTAATCTTTCTCAACATTTAGGAGCACTGGACCTGACCTTAATA 1020
DB 961 ACCATTTGAAATTAATCTTTCTCAACATTTAGGAGCACTGGACCTGACCTTAATA 1020
QY 1021 CGATTTTACAGCGTCACTAGTTAGATTTACATTAAGCATTAAGAGCCGTTCAAGC 1080
DB 1021 CGATTTTACAGCGTCACTAGTTAGATTTACATTAAGCATTAAGAGCCGTTCAAGC 1080
QY 1081 TATTTATACAAAGTTACAACTGATTTAGTGAATTTGATTTGATTTGATTTGAT 1140
DB 1081 TATTTATACAAAGTTACAACTGATTTAGTGAATTTGATTTGATTTGATTTGAT 1140
QY 1141 CCGGTTGATTAATTAATATGATTTAGTGAATTTGATTTGATTTGATTTGATTTG 1200
DB 1141 CCGGTTGATTAATTAATATGATTTAGTGAATTTGATTTGATTTGATTTGATTTG 1200
QY 1201 ACATATATACATTAATTTCTTACAGAAAAACCAACTTAAGAGAACTTAACATAT 1260
DB 1201 ACATATATACATTAATTTCTTACAGAAAAACCAACTTAAGAGAACTTAACATAT 1260
QY 1261 GGGTATGCTATACCTTTACAGATATGCTATGCTATGCTATGCTATGCTATGCT 1320
DB 1261 GGGTATGCTATACCTTTACAGATATGCTATGCTATGCTATGCTATGCTATGCT 1320
QY 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
DB 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380

US-09-502-426-1

Query Match	100.0%	Score 3202;	DB 19;	Length.6888;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 3202;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY	1	TTGCGGGATATATNTGTTGGTTCGGTTTGAGCTACAAATATAATTTTCGTTCTCGT	60
Db	1	ATGCGGGATATATNTGTTGGTTCGGTTTGAGCTACAAATATAATTTTCGTTCTCGT	60
QY	61	TATTCGTTCACAGATTTGAGTTTGCTTCAAATTTGATGTCACAGATTAATTAAT	120
Db	61	TATTCGTTCACAGATTTGAGTTTGCTTCAAATTTGATGTCACAGATTAATTAAT	120
QY	121	AAAAATCATTTAAATATTTACAAGTAATTAATATCTTTACATTGATGTTGTAACAA	180
Db	121	AAAAATCATTTAAATATTTACAAGTAATTAATATCTTTACATTGATGTTGTAACAA	180
QY	181	AATATCTATCTTGGTATATGAGAAAAATGAGAGTTGGAATTTATATAATTAAGGAA	240
Db	181	AATATCTATCTTGGTATATGAGAAAAATGAGAGTTGGAATTTATATAATTAAGGAA	240
QY	241	TAAATCATCTTGGTTGGTTGATTAACAGATTTTGTTGTTGCTTTGTTGTAATAGT	300
Db	241	TAAATCATCTTGGTTGGTTGATTAACAGATTTTGTTGTTGCTTTGTTGTAATAGT	300
QY	301	ATATGAGTAAATCAAAAAAGATATTGATGTAAGTGTAAACATATTTGTTATGACCCCA	360
Db	301	ATATGAGTAAATCAAAAAAGATATTGATGTAAGTGTAAACATATTTGTTATGACCCCA	360
QY	361	AAAAAAAAAAAAAAAACAACAAACAAACCCCCCGGATATAGTTTGGTTGTCGAGTT	420
Db	361	AAAAAAAAAAAAAAAACAACAAACAAACCCCCCGGATATAGTTTGGTTGTCGAGTT	420
QY	421	AGGTTTATTTGATCATATAATTCATGATCATTTCTTTGATTAAGTAAGATTTTCTTAC	480
Db	421	AGGTTTATTTGATCATATAATTCATGATCATTTCTTTGATTAAGTAAGATTTTCTTAC	480
QY	481	CAATTAATAATTTGCAATTCATATCTCTGATTTATTAATTAATACGAGTGTGAATATCC	540
Db	481	CAATTAATAATTTGCAATTCATATCTCTGATTTATTAATTAATACGAGTGTGAATATCC	540
QY	541	GTTTATGATCCTCCATCATCATGATTTATGATTTCTTGTCATCCAGCAAAATTAATACA	600
Db	541	GTTTATGATCCTCCATCATCATGATTTATGATTTCTTGTCATCCAGCAAAATTAATACA	600
QY	601	AGAGTATTGAGAAAAAACCCGAAATTAAGAAAAAGGAAAGAGTAGTACCCATGCGATG	660
Db	601	AGAGTATTGAGAAAAAACCCGAAATTAAGAAAAAGGAAAGAGTAGTACCCATGCGATG	660
QY	661	TGAATTAATTAACAAGAGATTAAGAGATGAACACAAAGGTTGGGAATATAGTCCCT	720
Db	661	TGAATTAATTAACAAGAGATTAAGAGATGAACACAAAGGTTGGGAATATAGTCCCT	720
QY	721	GCCAGCTTTCCTCACAATCATATATGACCCCTATTTGGATTTTCTGATATTCGTTAAA	780
Db	721	GCCAGCTTTCCTCACAATCATATATGACCCCTATTTGGATTTTCTGATATTCGTTAAA	780
QY	781	TTTTCGATTAACGATTTGAAAAAATATTTTATTTGTTAGCTGATCTCAATATTAATGTTCA	840
Db	781	TTTTCGATTAACGATTTGAAAAAATATTTTATTTGTTAGCTGATCTCAATATTAATGTTCA	840
QY	841	GGATTTTGCATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTTTGGTTCTGTTCT	900
Db	841	GGATTTTGCATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTTTGGTTCTGTTCT	900
QY	901	TAACTATATATTAATGCGGATATATGATTAACATGATATATACAAAAAATTTGTCGG	960
Db	901	TAACTATATATTAATGCGGATATATGATTAACATGATATATACAAAAAATTTGTCGG	960
QY	961	ACCAATTTTGAATAAATCTTTTCTCAAAACATTACGGGACACTGACCTGACCCTTAAATA	1020
Db	961	ACCAATTTTGAATAAATCTTTTCTCAAAACATTACGGGACACTGACCTGACCCTTAAATA	1020

QY	1021	CGATTTTACACGCGACAGTGTGGATTACTACGATAAAGCATTAAGACCGCTTCACG	1080
Db	1021	CGATTTTACACGCGTCACTAGTTGAGATTACTACGATTAAGCATTAAGACCGCTTCACG	1080
QY	1081	TATTTATACAAAGTTACAACTGCAATATAGCTTGAATTCCTTTAGAAAATTTTGGAAATTA	1140
Db	1081	TATTTATACAAAGTTACAACTGCAATATAGCTTGAATTCCTTTAGAAAATTTTGGAAATTA	1140
QY	1141	CCGGTGTGTATGTAAATATATAGATTAGTGGTAAACAAATATGTATATCATATATGTCGA	1200
Db	1141	CCGGTGTGTATGTAAATATATAGATTAGTGGTAAACAAATATGTATATCATATATGTCGA	1200
QY	1201	ACATATACATATATTCCTTACGAAAAAAACAACCTTAAGAGAGAGTTAAACATATCCATATAT	1260
Db	1201	ACATATACATATATTCCTTACGAAAAAAACAACCTTAAGAGAGAGTTAAACATATCCATATAT	1260
QY	1261	GGGTATGTATACCTTTCACGATGTATATCTATAGACACTAAAGATAGTATGTGATGTC	1320
Db	1261	GGGTATGTATACCTTTCACGATGTATATCTATAGACACTAAAGATAGTATGTGATGTC	1320
QY	1321	GATAAATGAATTTACACGCGGTGTAAATATATGGACCGTATGTATACGATCACTGCAG	1380
Db	1321	GATAAATGAATTTACACGCGGTGTAAATATATGGACCGTATGTATACGATCACTGCAG	1380
QY	1381	ATATCATTTCTGTTGGTGTCAACATTAATAAAACAAACAGAAAAAGAAAAACGATTTTT	1440
Db	1381	ATATCATTTCTGTTGGTGTCAACATTAATAAAACAAACAGAAAAAGAAAAACGATTTTT	1440
QY	1441	CTTGCGATTTCCATTCATATGATCTAAATATGCAATAGATCTTTGGCTTACAGTTTGCAGTCC	1500
Db	1441	CTTGCGATTTCCATTCATATGATCTAAATATGCAATAGATCTTTGGCTTACAGTTTGCAGTCC	1500
QY	1501	TCTACAAAGCGGTAAACATCTGCAAGCATTTAAATGCTTCTTAAATGCAATCTTAAACAT	1560
Db	1501	TCTACAAAGCGGTAAACATCTGCAAGCATTTAAATGCTTCTTAAATGCAATCTTAAACAT	1560
QY	1561	ATTTATTTGTTAGTTGGAATTTAATPAAGAGCACTGTAAACATTAATTAATTAATAG	1620
Db	1561	ATTTATTTGTTAGTTGGAATTTAATPAAGAGCACTGTAAACATTAATTAATTAATTAATAG	1620
QY	1621	ATATCTAGATGTGATTTTCCAAATTAACATCTTGGAGTTTAAACCTTAATCTTGGTTCT	1680
Db	1621	ATATCTAGATGTGATTTTCCAAATTAACATCTTGGAGTTTAAACCTTAATCTTGGTTCT	1680
QY	1681	TCCTACGGTATAAATTAATCATCGAGAGTAAAAAAGTTTGTATTTTCCGCGATGC	1740
Db	1681	TCCTACGGTATAAATTAATCATCGAGAGTAAAAAAGTTTGTATTTTCCGCGATGC	1740
QY	1741	ATGAAGATTAACCTATAGCTTAAATTTTTTGAATATGTAACCTTTACTCTATACATAT	1800
Db	1741	ATGAAGATTAACCTATAGCTTAAATTTTTTGAATATGTAACCTTTACTCTATACATAT	1800
QY	1801	AATTAACGGTATGTTTTTGTGTCCTAATATGACAGCTCTACAACTGATATGTCAAATTTTT	1860
Db	1801	AATTAACGGTATGTTTTTGTGTCCTAATATGACAGCTCTACAACTGATATGTCAAATTTTT	1860
QY	1861	TCCTCAAAATATTAATATGGAATTCATATGCTACTATCTAATATGAGAAACAGCTGAGTAT	1920
Db	1861	TCCTCAAAATATTAATATGGAATTCATATGCTACTATCTAATATGAGAAACAGCTGAGTAT	1920
QY	1921	ACATTTTAATTTAAGCAAAATTTTGAATAAGTATATATTTCTAACATATATTAATA	1980
Db	1921	ACATTTTAATTTAAGCAAAATTTTGAATAAGTATATATTTCTAACATATATTAATA	1980
QY	1981	AATATGATGCTAATAATGATATTTCCATATGTTCTTAAATATTTTTTTTATATTAAGTAA	2040
Db	1981	AATATGATGCTAATAATGATATTTCCATATGTTCTTAAATATTTTTTTTATATTAAGTAA	2040
QY	2041	TAAATACATTAATGACCAATATATATGTTGGGAAATTCAAATATCTCCATTAATATTTTGG	2100
Db	2041	TAAATACATTAATGACCAATATATATGTTGGGAAATTCAAATATCTCCATTAATATTTTGG	2100

QY	2101	AAATCTCAAAATTAATTAATTTAGTCAATTAACAATGCATAGAAAGTTCCAAAAAATTT	2160
Db	2101	AAATCTCAAAATTAATTAATTTAGTCAATTAACAATGCATAGAAAGTTCCAAAAAATTT	2160
QY	2161	TTGTAAACAAACTCCAAATTTTTTTTTTTTATAGCAACAAGAAATTAACGATAGAAA	2220
Db	2161	TTGTAAACAAACTCCAAATTTTTTTTTTTTATAGCAACAAGAAATTAACGATAGAAA	2220
QY	2221	CTATTTTGTGTGGAAATGGAAAGTAGTAATATACATTAAGCAAAATTTTAAAAATTTATATA	2280
Db	2221	CTATTTTGTGTGGAAATGGAAAGTAGTAATATACATTAAGCAAAATTTTAAAAATTTATATA	2280
QY	2281	AGCCTTACGCGGTCAAAATATGTATATCTAGTAGCTGTATTAATATATCATGCTGTGCAT	2340
Db	2281	AGCCTTACGCGGTCAAAATATGTATATCTAGTAGCTGTATTAATATATCATGCTGTGCAT	2340
QY	2341	TCGAATTTGGGCAACAATGAAGAAAGGAATTTAAATATTAATCTTTAAAAATTAATAAAT	2400
Db	2341	TCGAATTTGGGCAACAATGAAGAAAGGAATTTAAATATTAATCTTTAAAAATTAATAAAT	2400
QY	2401	TTGAGTAATATGTGTTTCTGACATATTGAGGGCAAAAAAAGACAAATGCCAAAAGTCTAC	2460
Db	2401	TTGAGTAATATGTGTTTCTGACATATTGAGGGCAAAAAAAGACAAATGCCAAAAGTCTAC	2460
QY	2461	GGGTTTGACTCTCCAGTTCGGTAATATCTATATACCTGTCTTGACCCGACGCTGTG	2520
Db	2461	GGGTTTGACTCTCCAGTTCGGTAATATCTATATACCTGTCTTGACCCGACGCTGTG	2520
QY	2521	TAGGGTCCCTTGTGACATTTTACAGTTCACCTGTCTACCCCTACTGCTGAGCCACCTTTCCCA	2580
Db	2521	TAGGGTCCCTTGTGACATTTTACAGTTCACCTGTCTACCCCTACTGCTGAGCCACCTTTCCCA	2580
QY	2581	TATCCTAAGGGTAATTTTGGAAATCCCAATTTTAACGATTTGAGACCGCTACCGGACTTC	2640
Db	2581	TATCCTAAGGGTAATTTTGGAAATCCCAATTTTAACGATTTGAGACCGCTACCGGACTTC	2640
QY	2641	TGGGATTTGCTGCTGGAGCAATTAACAAAATTTATAGCAGAAATGGGTTATTATATTAA	2700
Db	2641	TGGGATTTGCTGCTGGAGCAATTAACAAAATTTATAGCAGAAATGGGTTATTATATTAA	2700
QY	2701	AACTACAACTTGATCAGATTAATTTTCATTAACACTTTTACATAGGATTCGTAGCATCT	2760
Db	2701	AACTACAACTTGATCAGATTAATTTTCATTAACACTTTTACATAGGATTCGTAGCATCT	2760
QY	2761	ATCTAATGACTTTTTTTTTTCTACACGAGTGGAAGTAATAGTACTATTAGCCAGAG	2820
Db	2761	ATCTAATGACTTTTTTTTTTCTACACGAGTGGAAGTAATAGTACTATTAGCCAGAG	2820
QY	2821	ACAATGTGTTATAGATATATCATTATATCCATATATATATATATTAATTAATTAACCTGTAAA	2880
Db	2821	ACAATGTGTTATAGATATATCATTATATCCATATATATATATATTAATTAATTAACCTGTAAA	2880
QY	2881	CTATTTGACATGCACTTTCTGCAACTTTCTTTTATTTAATTAAGTTTAAATTAATA	2940
Db	2881	CTATTTGACATGCACTTTCTGCAACTTTCTTTTATTTAATTAAGTTTAAATTAATA	2940
QY	2941	AAAGTAATTAAGGAGCAATAACGAGCAACAAGTAATGACAGGAGCAACAAAGGCA	3000
Db	2941	AAAGTAATTAAGGAGCAATAACGAGCAACAAGTAATGACAGGAGCAACAAAGGCA	3000
QY	3001	TGAAGCTCATTTGTTAGTTTAAAGCTTAATAGAAAGTTTATTAATTTAATGACATG	3060
Db	3001	TGAAGCTCATTTGTTAGTTTAAAGCTTAATAGAAAGTTTATTAATTTAATGACATG	3060
QY	3061	ATAACCAATTAATTTTCTGACTTCTTTAAACCCCTCTTACAAACAGAGCTCCCTTT	3120
Db	3061	ATAACCAATTAATTTTCTGACTTCTTTAAACCCCTCTTACAAACAGAGCTCCCTTT	3120
QY	3121	TCAGTGAAGTCCGATTTCCCAATCTTAAAGCAAAAGCAATTAAGAAAGAAAGTGA	3180
Db	3121	TCAGTGAAGTCCGATTTCCCAATCTTAAAGCAAAAGCAATTAAGAAAGAAAGTGA	3180
QY	3181	GAAGAGAGAGAAACTAGCTCC	3202

Db

3181 GAGAGAGAGAGAACTACTCTC 3202

|||||

RESULT 3

US-09-692-412-94/C

; Sequence 94, Application US/09692412

; GENERAL INFORMATION:

; APPLICANT: Bush, David F.

; APPLICANT: Rounsley, Steven D.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof

; FILE REFERENCE: 38-10(15493)C

; CURRENT APPLICATION NUMBER: US/09/692,412

; CURRENT FILING DATE: 2000-10-24

; NUMBER OF SEQ ID NOS: 124

; SEQ ID NO 94

; LENGTH: 84196

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-692-412-94

Query Match	97.9%;	Score 3134.2;	DB 27;	Length 84196;
Best Local Similarity	99.5%;	Pred. No. 0;		
Matches 3186;	Conservative	0;	Mismatches 13;	Indels 4;
				Gaps 4

QY	1	ATGTGGGTAATTAATTTGGGTTCGGTTTGACCTACAAATPAATTAATTCGTTCGTGT	60
Db	32199	ATGTGGGTAATTAATTTGGGTTCGGTTTGACCTACAAATPAATTAATTCGTTCGTGT	32141
QY	61	TATTCGTTCACATGATTTGAGTTTGCTCAATTTGGATTCGCAAGTAATTAATTAAT	120
Db	32139	TATTCGTTCACATGATTTGAGTTTGCTCAATTTGGATTCGCAAGTAATTAATTAAT	32068
QY	121	AAAATTCATTTAAAAATTTTACAAAGTAATTAATTCCTTACATTGTATTTGTTATACAA	180
Db	32079	AAAATTCATTTAAAAATTTTACAAAGTAATTAATTCCTTACATTGTATTTGTTATACAA	32070
QY	181	AATATCTATCTTTGGTATATGAGAAAATTTGAGTTTGGAAATTTATTAATTAAGGAA	240
Db	32019	AATATCTATCTTTGGTATATGAGAAAATTTGAGTTTGGAAATTTATTAATTAAGGAA	31966
QY	241	TATATGATTCCTATTTGGTTGGATTCACAGTTAACTTTTGTGTTCTTTTGTATATGT	300
Db	31959	TATATGATTCCTATTTGGTTGGATTCACAGTTAACTTTTGTGTTCTTTTGTATATGT	31900
QY	301	ATATGAGTAAATCAAAAAGAGATTTGATTTGAAGTAAACAATTTGCTTATGACCCCA	360
Db	31899	ATATGAGTAAATCAAAAAGAGATTTGATTTGAAGTAAACAATTTGCTTATGACCCCA	31844
QY	361	AAAAAAAAAAAAAAAACAAAACAAACCCCCCGATPAAGTTTGTGTTCTGTGATT	420
Db	31839	AAAAAAAAAAAAAAAACAAAACAAACACA-NCATTTTTCGATPAAGTTTGTGTTCTGTGATT	31788
QY	421	AGGTTATTTGATCATTAATTCATCATCATTTCTTGGATACGATAGAAATTTCTTAC	480
Db	31780	AGGTTATTTGATCATTAATTCATCATCATTTCTTGGATACGATAGAAATTTCTTAC	31722
QY	481	CAATTAATAATTTCCGAATTCATATCTCTTGATTTATTAATTAATACAGAGTGTGAATATCC	540
Db	31720	CAATTAATAATTTCCGAATTCATATCTCTTGATTTATTAATTAATACAGAGTGTGAATATCC	31666
QY	541	GTTATATGATCAGTCCCATCATGATTTGATTTCTTGCGATACCGCAAAATTAATTAACA	600
Db	31660	GTTATATGATCAGTCCCATCATGATTTGATTTCTTGCGATACCGCAAAATTAATTAACA	31605
QY	601	AGAGTATTTGAGAAAAAAGGAAAAATTAAGAAAGGAAAGAGTAGTACCCTGAGTAGT	660
Db	31600	AGAGTATTTGAGAAAAAAGGAAAAATTAAGAAAGGAAAGAGTAGTACCCTGAGTAGT	31544
QY	661	TGATATATTTCAAAAGAGATAAGAGTGTGCAACCAAAAGTTGTGGAATTAATGGTCCCT	720
Db	31540	TGATATATTTCAAAAGAGATAAGAGTGTGCAACCAAAAGTTGTGGAATTAATGGTCCCT	31481

QY 721 GCCAGCTTCTCACAATCAATATGCAACCCATTTGGATTTCTGGATATTCGTTAAA 780
Db 31480 GCCAGCTTCTCACAATCAATATGCAACCCATTTGGATTTCTGGATATTCGTTAAA 31421
QY 781 TTTCGGCAATGAGATGGGAAAAATATTTATTTGTGTAGCTGATCTCAATATATGTTCCA 840
Db 31420 TTTCGGCAATGAGATGGGAAAAATATTTATTTGTGTAGCTGATCTCAATATATGTTCCA 31361
QY 841 GGTATTTGCAATATCTCTGTGTTTAAGCAATTTTGTCTTTCTTTTGTGTTCTCT 900
Db 31360 GGTATTTGCAATATCTCTGTGTTTAAGCAATTTTGTCTTTCTTTTGTGTTCTCT 31301
QY 901 TAACATATATATATGCGGATATATGATPACAAATGATATATACAAAACAATGTCGGG 960
Db 31300 TAACATATATATATGCGGATATATGATPACAAATGATATATACAAAACAATGTCGGG 31241
QY 961 ACCATTTGSAATTAACCTTTTCTCAACATATAGGGGACATGAGCTGACCCCTTAATA 1020
Db 31240 ACCATTTGSAATTAACCTTTTCTCAACATATAGGGGACATGAGCTGACCCCTTAATA 31182
QY 1021 CGATTTTACAGCGCTAGTATGATTTAGTATAGCATAAAGCATPAAAGACCCGTTCAAGC 1080
Db 31181 CGATTTTACAGCGCTAGTATGATTTAGTATAGCATAAAGCATPAAAGACCCGTTCAAGC 31122
QY 1081 TATTTATACAAAGTTACAAACCTGATATAGCTTGAATTCCTTTAGAAAATTTTGGAAATTA 1140
Db 31121 TATTTATACAAAGTTACAAACCTGATATAGCTTGAATTCCTTTAGAAAATTTTGGAAATTA 31062
QY 1141 CCGGCTGTATGTAATATATGATTTAGTAAACAAATATGTAATCAATATATGTCGA 1200
Db 31061 CCGGCTGTATGTAATATATGATTTAGTAAACAAATATGTAATCAATATATGTCGA 31002
QY 1201 ACATATACATATATCTTTACAGAAAAACAACCTTAAGAGAGTTAATCATATCCATATAT 1260
Db 31001 ACATATACATATATCTTTACAGAAAAACAACCTTAAGAGAGTTAATCATATCCATATAT 30942
QY 1261 GGGATGCTATACCTTTACAGATGCTATATCTAGAGCTPAAAGATAGTATGATGTC 1320
Db 30941 GGGATGCTATACCTTTACAGATGCTATATCTAGAGCTPAAAGATAGTATGATGTC 30882
QY 1321 GATTAATGAAATTCACACGGGTGTAATATATGGAACGCTATGTAACGATCTCAGAA 1380
Db 30881 GATTAATGAAATTCACACGGGTGTAATATATGGAACGCTATGTAACGATCTCAGAA 30822
QY 1381 ATATCATCTCTGGTGTCAACAATAAAAAACAAGAAAAAGAAAAAGAAAAAGCAATTTT 1440
Db 30821 ATATCATCTCTGGTGTCAACAATAAAAAACAAGAAAAAGAAAAAGAAAAAGCAATTTT 30762
QY 1441 CTGGAATTCATTCATGATCTAAATATGATCTTTTGGGTTACAGTTTCGAAGTCC 1500
Db 30761 CTGGAATTCATTCATGATCTAAATATGATCTTTTGGGTTACAGTTTCGAAGTCC 30702
QY 1501 TCTACAAGCGTAAACCATGCAATATTAATTTGCTTTCTTAATGACATCTTAACAT 1560
Db 30701 TCTACAAGCGTAAACCATGCAATATTAATTTGCTTTCTTAATGACATCTTAACAT 30642
QY 1561 ATTATATGTAAGTGAATTAATAAGAGGAACTTGAACATTAACATATATATATAG 1620
Db 30641 ATTATATGTAAGTGAATTAATAAGAGGAACTTGAACATTAACATATATATATAG 30582
QY 1621 ATACTAGATGATTTTCCAAATACATCTTTGGATGTTTAAACCTTAATCTTGTCT 1680
Db 30581 ATACTAGATGATTTTCCAAATACATCTTTGGATGTTTAAACCTTAATCTTGTCT 30522
QY 1681 TCCAGGATTAATATATTAATCAATGAGTAAAGAAAAAGTTTGTCTATTTTGGCGAGTC 1740
Db 30521 TCCAGGATTAATATATTAATCAATGAGTAAAGAAAAAGTTTGTCTATTTTGGCGAGTC 30463
QY 1741 ATGAAGGATTAACCTAATGATTAATTTTGAAGATGTAACCCCTTTACTCATATAT 1800
Db 30462 ATGAAGGATTAACCTAATGATTAATTTTGAAGATGTAACCCCTTTACTCATATAT 30403

QY 1801 AATTACCGTATGTTTTGTTGCCAATATGAGACGCTCTACACGCTGTGATGATCAATTTT 1860
Db 30402 AATTACCGTATGTTTTGTTGCCAATATATGAGACGCTCTACACGCTGTGATGATCAATTTT 30343
QY 1861 TCTGCAATATATTAATTAAGAAATTCATATGCTACATATGAAATGAAAGAACCTGATAT 1920
Db 30342 TCTGCAATATATTAATTAAGAAATTCATATGCTACATATGAAATGAAAGAACCTGATAT 30283
QY 1921 ACATTTTAATTAAGACAAATTTTGAAGAAAGTATATTTCTTAACAATATATATTA 1980
Db 30282 ACATTTTAATTAAGACAAATTTTGAAGAAAGTATATTTCTTAACAATATATATTA 30223
QY 1981 AATATGATGCTTAATATGATATTCCTATGTTCTTAATTAATTTTATATATATATTA 2040
Db 30222 AATATGATGCTTAATATGATATTCCTATGTTCTTAATTAATTTTATATATATTA 30163
QY 2041 TAAATACATTAATGAAACCAATATGTTGGAATTCAAATATCCCATATATTTTGTG 2100
Db 30162 TAAATACATTAATGAAACCAATATGTTGGAATTCAAATATCCCATATATTTTGTG 30103
QY 2101 AATCTACAAATTAATTAATATTTAGTCAATTAACAATGATGATGAAAGTTCCAAAAAAT 2160
Db 30102 AATCTACAAATTAATTAATATTTAGTCAATTAACAATGATGATGAAAGTTCCAAAAAAT 30043
QY 2161 TTGTTAAACGAACCTCCAAATTTTTTTTTTATGCAACAGAAATATACAGATGAAAA 2220
Db 30042 TTGTTAAACGAACCTCCAAATTTTTTTTTTATGCAACAGAAATATACAGATGAAAA 22983
QY 2221 CTATTTGTTGGAATGGAAGTATATATATATATATATATATATATATATATATAT 2280
Db 22982 CTATTTGTTGGAATGGAAGTATATATATATATATATATATATATATATATATATAT 22923
QY 2281 AGCCTATACGCGCTCAAAAGTATGTTATCTAGTAGTGAATTAATATATGATGTCGAT 2340
Db 22922 AGCCTATACGCGCTCAAAAGTATGTTATCTAGTAGTGAATTAATATATGATGTCGAT 22963
QY 2341 TCAGAAATGGGACAAACAATGAAAAAGAAATTAATTAATCTTTAAATTAATTAAT 2400
Db 22962 TCAGAAATGGGACAAACAATGAAAAAGAAATTAATTAATCTTTAAATTAATTAAT 22983
QY 2401 TTGAGTAAATGTTTCTGCTGCTATGAGGGGCAAAAAAGAACATCCAAAAAGTCTAC 2460
Db 22982 TTGAGTAAATGTTTCTGCTGCTATGAGGGGCAAAAAAGAACATCCAAAAAGTCTAC 22943
QY 2461 GGGTTGACGTGCGAGTCCGTAATTAATCTAATACTGCTTTGACCGGACGCTGCTG 2520
Db 22942 GGGTTGACGTGCGAGTCCGTAATTAATCTAATACTGCTTTGACCGGACGCTGCTG 22963
QY 2521 TAGGGGCTCTTCTACATTTTCACTGTTTCAACCCCTACTGCTGAGCCACCCCTTTCCCA 2580
Db 22962 TAGGGGCTCTTCTACATTTTCACTGTTTCAACCCCTACTGCTGAGCCACCCCTTTCCCA 22923
QY 2581 TATCCTAAGGTAATTTTGAAGATCCCAATTAACGATGAGACGCTACCGGACTCC 2640
Db 22922 TATCCTAAGGTAATTTTGAAGATCCCAATTTAAACGATGAGACGCTACCGGACTCC 22953
QY 2641 TGGGATCTGCTGAGCATTTATCAAAAAATTAATGACAGAAATGGGTTTATTAATTA 2700
Db 22952 TGGGATCTGCTGAGCATTTATCAAAAAATTAATGACAGAAATGGGTTTATTAATTA 22953
QY 2701 AACTCACAACCTGATCAATTAATTTCAATTAACACTTTTACGATGATTTGCTACGATCT 2760
Db 22952 AACTCACAACCTGATCAATTAATTTCAATTAACACTTTTACGATGATTTGCTACGATCT 22943
QY 2761 ATCTAATGACTTTTTTTTTTCTACACGAGTGAGTAAAGTATATGATATATAGCAGAG 2820
Db 22942 ATCTAATGACTTTTTTTTTTCTACACGAGTGAGTAAAGTATATGATATATAGCAGAG 22983
QY 2821 ACAATGATTAATGATATATCCATTAATCCATGATATTAATTAATTAATAGCTTAA 2880
Db 22982 ACAATGATTAATGATATATCCATTAATCCATGATATTAATTAATTAATAGCTTAA 22923
QY 2881 CTATTTACGATCCGACCTTTCTGCAACTTTTGTTTTATTTAAGAGTTTAAATTA 2940

[illegible]

RESULT 4
US-09-803-736-1074/C
Sequence 1074, Application US/09803736
GENERAL INFORMATION:
APPLICANT: Bush, David F.
APPLICANT: Levin, Irena M.
APPLICANT: Norris, Susan R.
APPLICANT: Rounsley, Steven D.
APPLICANT: Miegand, Roger C.
TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
FILE REFERENCE: 38-10(15493)D
CURRENT APPLICATION NUMBER: US/09/803.736
CURRENT FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 09/534,859
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPU01 38-10
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ. ID NOS: 1582
SEQ. ID NO 1074
LENGTH: 84196
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-803-736-1074

Query Match	97.9%	Score 3134.2	DB 31	Length 84196
Best Local Similarity	99.5%	Pred. No. 0		
Matches 3186	Conservative	0	Mismatches 13	Indels 4
				Gaps 4
QY	1	ATGTGGGATTTATATATGTTGGGTTGCGGTTGAGCTCAATATPAAATTTGGTGTGGT	60	
Db	32199	ATGTGGGATTTATATATGTTGGGTTGCGGTTGAGCTCAATATPAAATTTGGTGTGGT	32140	
QY	61	TATTCGTTACATGATTTGAGTTGGTCTCAATTTGGATTCCAAGATTAATTAAT	120	
Db	32139	TATTCGTTACATGATTTGAGTTGGTCTCAATTTGGATTCCAAGATTAATTAAT	32080	
QY	121	AAAATTCATTTAAATAATTTACAGTAATTAATTAATCTTTACATTGTATTATACAA	180	
Db	32079	AAAATTCATTTAAATAATTTACAGTAATTAATTAATCTTTACATTGTATTATACAA	32020	
QY	181	AAATATCATCTTGGATATATGAGAAATATPAGAGTTGGATTAATTAATTAAGGAA	240	
Db	32019	AAATATCATCTTGGATATATGAGAAATATPAGAGTTGGATTAATTAATTAAGGAA	31960	
QY	241	TAAATTCATTCATTTGGTGGATTACACAGTTAACTTTTGGTTCTTTGTATATGT	300	
Db	31959	TAAATTCATTCATTTGGTGGATTACACAGTTAACTTTTGGTTCTTTGTATATGT	31900	
QY	301	ATATGAGTAAATCAAAAAGATTTGATTTGAAGTGAATTAATTTGGTATGACCCCA	360	

Db	31899	ATATGAGTAATTCACAAAAGACAAATGGATTGAAGTGAATCAATATTTCTGTTATGACACCCCC	31840
Qy	361	AAAAAAAAAAAAAAAAACAACAAACAAACCCCCCCCGACATATAGTTTGGTCTCGATT	420
Db	31899	CAAAAAAAAAAAAAAAAAACAACAAACAA - ACAATTTTCGATATAGTTTGGTCTCGATT	31761
Qy	421	AGGTTTATTGATCATATTTATTCATGCAATTTCTTGATTACTATGAAGATTTCTATC	480
Db	31780	AGGTTTATTGATCATATTTATTCATGCAATTTCTTGATTACTATGAAGATTTCTATC	31721
Qy	481	CAATTAATTAATTCGAATTCATATCTCGATTATTAATTAATTAATTAATTAATTAATTC	540
Db	31720	CAATTAATTAATTCGAATTCATATCTCGATTATTAATTAATTAATTAATTAATTAATTC	31661
Qy	541	GTTTATGATCAGTCACATCATGATTTATGATTTCTGTGTAATTCACGAAATTTATTAACA	600
Db	31660	GTTTATGATCAGTCACATCATGATTTATGATTTCTGTGTAATTCACGAAATTTATTAACA	31601
Qy	601	AGAGATTTGACAAAACAAACCCGAAATTAAGAAAAGGAAAGGATAGTACCCATGAGATAG	660
Db	31600	AGAGATTTGACAAAACAAACCCGAAATTAAGAAAAGGAAAGGATAGTACCCATGAGATAG	31541
Qy	661	TGAATTAATTAACAAGAGAAATTAAGATGACAAACCAAAAGGTTGTGAATTAATGTCCT	720
Db	31540	TGAATTAATTAACAAGAGAAATTAAGATGACAAACCAAAAGGTTGTGAATTAATGTCCT	31481
Qy	721	GCCAGCTTCTCTGACATCAATATGACACCTATTTGGATTTCTGGAATTCGTTAAA	780
Db	31480	GCCAGCTTCTCTGACATCAATATGACACCTATTTGGATTTCTGGAATTCGTTAAA	31421
Qy	781	TTTTCGATTAACGATTTGGAATAATTTTATTTGTAAGCATCCTCAATATTAATGTTCCA	840
Db	31420	TTTTCGATTAACGATTTGGAATAATTTTATTTGTAAGCATCCTCAATATTAATGTTCCA	31361
Qy	841	GGATTTTGCATTAATCTTCTGTTTAAAGCATATTTTCTCTTTCTTTTGGTGGTCTCT	900
Db	31360	GGATTTTGCATTAATCTTCTGTTTAAAGCATATTTTCTCTTTCTTTTGGTGGTCTCT	31301
Qy	901	TAACTATATATTAACGGGATATATGATTAATGATATATATATATATATATATATATG	960
Db	31300	TAACTATATATTAACGGGATATATGATTAATGATTAATGATTAATATATATATATATATG	31241
Qy	961	ACCAATTTTGAATTAACCTTTTCTCAAAACATTAAGGACATGAGCTCGACCTTAAATA	1020
Db	31240	ACCAATTTTGAATTAACCTTTTCTCAAAACATTAAGGACATGAGCTCGACCTTAAATA	31182
Qy	1021	CGATTTTACAGCGCTCACTAGTTAGATTTACTAGCATTAAGCATTAAGGACCCGTTCAAGC	1080
Db	31181	CGATTTTACAGCGCTCACTAGTTAGATTTACTAGCATTAAGCATTAAGGACCCGTTCAAGC	31122
Qy	1081	TATTTATACAAAGTTACAACTGAATATACCTGGAATCCTTTAGAAAATTTTGGAAATTA	1140
Db	31121	TATTTATACAAAGTTACAACTGAATATACCTGGAATCCTTTAGAAAATTTTGGAAATTA	31062
Qy	1141	CCGGTTGTTATGTAATATATAGATTTAGTGTAAACAATATGTTAATCAATTAAGTGTC	1200
Db	31061	CCGGTTGTTATGTAATATATAGATTTAGTGTAAACAATATGTTAATCAATTAAGTGTC	31002
Qy	1201	ACATATTCATTAATCCTTACAGAAAACAAACTTTAAGAAAGTTTAACTATCCATATAT	1260
Db	31001	ACATATTCATTAATCCTTACAGAAAACAAACTTTAAGAAAGTTTAACTATCCATATAT	30942
Qy	1261	GGGTATGCTATACCTTACAGTATGCTATATAGAGCTTAAGAAATAGTTATGTATGTC	1320
Db	30941	GGGTATGCTATACCTTACAGTATGCTATATAGAGCTTAAGAAATAGTTATGTATGTC	30882
Qy	1321	GATTAATGAATTCACACGCGGTGCTAATTAATTAAGGACCGTATGTCAGATCTGCA	1380
Db	30881	GATTAATGAATTCACACGCGGTGCTAATTAATTAAGGACCGTATGTCAGATCTGCA	30822
Qy	1381	ATATCATCTTGTGGTCAACATTAATTAACAAACAAAGAAAGAAAGAAACGATTTT	1440

Best Local Similarity 98.6%; Pred. No. 1.6e-128;
Matches 983; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

```

QY 2208 AACAGATGAAAGATATTTTGTGGAGATGAGATGATATATCATTAACCAATTTT 2267
    |||
Db 1 AACAGATGAAAGATATTTTGTGGAGATGAGATGATATATCATTAACCAATTTT 60
QY 2268 AAAAAATATATAGCTATACGCGCTCAAGATATGTATCTAGTAGTATTAATA 2327
    |||
Db 61 AAAAAATATATAGCTATACGCGCTCAAGATATGTATCTAGTAGTATTAATA 120
QY 2328 TGCATGTGCGATTTAGAAATGGGACAAACATGAAACGGAATTAATTAACCTTAA 2387
    |||
Db 121 TGCATGTGCGATTTAGAAATGGGACAAACATGAAACGGAATTAATTAACCTTAA 180
QY 2388 AATAATATAAATTTGAGTAATGTGTTTCTGACTATGAGGGCAAAAAAGACAT 2447
    |||
Db 181 AATAATATAAATTTGAGTAATGTGTTTCTGACTATGAGGGCAAAAAAGACAT 240
QY 2448 GCCAAAGCTACGGGTTTGAAGTGTGCTGATTAATTAATTAATCTGCTTTGA 2507
    |||
Db 241 GCCAAAGCTACGGGTTTGAAGTGTGCTGATTAATTAATTAATCTGCTTTGA 300
QY 2508 CCGACAGCTCGTGTAGGGGCTCTTGACATTTTCACTGTTACCCCTAGCTGAGCC 2567
    |||
Db 301 CCGACAGCTCGTGTAGGGGCTCTTGACATTTTCACTGTTACCCCTAGCTGAGCC 360
QY 2568 CACCCCTTTCCCATTCCTAAGGGTAATTTTGGAAATCCCATTTAAACCGATGAGAC 2627
    |||
Db 361 CACCCCTTTCCCATTCCTAAGGGTAATTTTGGAAATCCCATTTAAACCGATGAGAC 420
QY 2628 GTACGGAGCTTCCCTGGATTCCTGCTGAGCATTTATCAAAATTTATACAGAAATGGT 2687
    |||
Db 421 GTACGGAGCTTCCCTGGATTCCTGCTGAGCATTTATCAAAATTTATACAGAAATGGT 480
QY 2688 TTATTAATTTAAAACTCACAACCTGATCAGATTAATTAATTAACCTTTAGCAGG 2747
    |||
Db 481 TTATTAATTTAAAACTCACAACCTGATCAGATTAATTAATTAACCTTTAGCAGG 540
QY 2748 ATTGCTAGCATCTATCTATATGACTTTTCTTCCACCGGGTGTGAAGTATATAGA 2807
    |||
Db 541 ATTGCTAGCATCTATCTATATGACTTTTCTTCCACCGGGTGTGAAGTATATAGA 600
QY 2808 CTATTAGCCAGAGACATTTGATATATATATCCATTATCCATGATATTTATGATATA 2867
    |||
Db 601 CTATTAGCCAGAGACATTTGATATATATATCCATTATCCATGATATTTATGATATA 660
QY 2868 AATAGCTGTAACTATTTCAGCATGCGAGCTTTGCAACTTTGTTTAAATTTAGA 2927
    |||
Db 661 AATAGCTGTAACTATTTCAGCATGCGAGCTTTGCAACTTTGTTTAAATTTAGA 720
QY 2928 GTTTAATTAATTAAGTATTAAGAGCATTAACGCAACAAAAGTATGAACAGCA 2987
    |||
Db 721 GTTTAATTAATTAAGTATTAAGAGCATTAACGCAACAAAAGTATGAACAGCA 780
QY 2988 GAAACAAAAGCCATGAAGCTCATTTGTTA-GTTAAGCTTAATAGAGATTTTATTA 3046
    |||
Db 781 GAAACAAAAGCCATGAAGCTCATTTGTTAAGCTTAATAGAGATTTTATTA 840
QY 3047 TTTTATATGAC-GATATATCAATTAATTTCTGACTTCTTTAAACCCCTCTTACAA 3105
    |||
Db 841 TTTTATATGACGATATATCAATTAATTTCTGACTTCTTTAAACCCCTCTTACAA 900
QY 3106 CAGAGCTCCCTTTTTCAGTGAAGTCCGATCCCAACTTAAACAAAGCATTAAGA 3165
    |||
Db 901 CAGAGCTCCCTTTTTCAGTGAAGTCCGATCCCAACTTAAACAAAGCATTAAGA 960
QY 3166 AGAGAAAGTGAAGTGAAGAGAGAGAAAGTACTGCC 3202
    |||
Db 961 AGAGAAAGTGAAGTGAAGAGAGAGAAAGTACTGCC 997

```

RESULT 6

```

PCT-US99-22853B-3486
; Sequence 3486, Application PC/TUS9922853B
; GENERAL INFORMATION:
; APPLICANT: Ceres, Inc.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
; FILE REFERENCE: Polypeptides Encoded Thereby
; CURRENT APPLICATION NUMBER: 2750-0569F(PC)
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 3938
; SOFTWARE: MS Word 97
; SEQ ID NO 3486
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: LOCATION 1..826, Ceres Seq. ID 1592085
; NAME/KEY: UNSURE
; LOCATION: (1)..(826)
; OTHER INFORMATION: any n = any nucleic acid, unknown or other
PCT-US99-22853B-3486

```

Query Match 3.9%; Score 125; DB 1; Length 826;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3078 TGACTTCTTAAACCCCTCTTACAAAGAGCTCCCTTTTCTAGTAGAATCCGATT 3137
    |||
Db 1 TGACTTCTTAAACCCCTCTTACAAAGAGAGCTCCCTTTTCTAGTAGAATCCGATT 60
QY 3138 CCCAATCTTAAAGACAAAGCCATTAGAAAGAGAAAGTGAAGAGAGAGAGAACTA 3197
    |||
Db 61 CCCAATCTTAAAGACAAAGCCATTAGAAAGAGAAAGTGAAGAGAGAGAGAACTA 120
QY 3198 GCTCC 3202
    |||
Db 121 GCTCC 125

```

```

RESULT 7
US-09-692-412-94
; Sequence 94, Application US/09692412
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)C
; CURRENT APPLICATION NUMBER: US/09/692,412
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 124
; SEQ ID NO 94
; LENGTH: 84196
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-692-412-94

```

Query Match 3.7%; Score 118; DB 27; Length 84196;
Best Local Similarity 83.2%; Pred. No. 3.7e-07;
Matches 158; Conservative 0; Mismatches 30; Indels 2; Gaps 2;

```

QY 763 TCTGATATTCGTTAAATTTGCGATTAAGATGTAAGAAATATTTATTTGTTACTGA 822
    |||
Db 33161 TATAGAAATTCGTTAAATTTGCGATTAAGATGTAAGAAATATTTATTTGTTACTGA 33220
QY 823 TCTCAATATTAATGTTCCAGGATTTGCGATTAATCTTCTTTAAAGCATA-TTTTGTCTT 881
    |||
Db 33221 TCTCAATATTAATTTCCAGGATTTT-CATTAATCTTCTTTAAAGATATCAATTTCTAT 33279
QY 882 CTTTGTGTTCTTCTTCTTACATATATATATCGGGATATATGATATCAATGATATAT 941
    |||
Db 33280 CTTTGTGTTCTTCTTCTTACATATATATATGCGGATATATGAAACAAAGTCTTAG 33339

```

QY 942 CACAAACAA 951
 DB 33340 TTACAAAAA 33349

RESULT 8

US-09-803-736-1074
 ; Sequence 1074, Application US/09803736
 ; GENERAL INFORMATION:
 ; APPLICANT: Bush, David F.
 ; APPLICANT: Levin, Irena M.
 ; APPLICANT: Norris, Susan R.
 ; APPLICANT: Rounsley, Steven D.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
 ; FILE REFERENCE: 38-10(15493)D
 ; CURRENT APPLICATION NUMBER: US/09/803/736
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/534,859
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
 ; PRIOR FILING DATE: 2000-10-20
 ; NUMBER OF SEQ ID NOS: 1582
 ; SEQ ID NO 1074
 ; LENGTH: 84196
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-803-736-1074

Query Match 3.7%; Score 118; DB 31; Length 84196;
 Best Local Similarity 83.2%; Pred. No. 3.7e-07;
 Matches 158; Conservative 0; Mismatches 30; Indels 2; Gaps 2;

QY 763 TCTGATATTCGTTAAATTTGGATACGATGTGAAATATTTATTTGTTAGCTGA 822
 DB 33161 TATAGAAATTCGTTAAATTTGGATACGATGTGAAATATTTATTTGTTAGCTGA 33220
 QY 823 TCTCAATATTTATGTTCCAGGATTTGATATCTCTGTTTAAAGCATA-TTTTGTCTT 881
 DB 33221 TCTCAATATTTATGTTCCAGGATTTGATATCTCTGTTTAAAGCATA-TTTTGTCTT 33279
 QY 882 CTTTTCGTTCTGTTCTTACATATTTATTCGCGATATATGATTAACAAATGATAT 941
 DB 33280 CTTTTCGTTCTGTTCTTACATATTTATTCGCGATATATGATTAACAAATGATAT 33339
 QY 942 CACAAACAA 951
 DB 33340 TTACAAAAA 33349

RESULT 9

US-60-207-458-33653
 ; Sequence 33653, Application US/60207458
 ; GENERAL INFORMATION:
 ; APPLICANT: Abad, Mark S.
 ; APPLICANT: Conner, Timothy W.
 ; APPLICANT: Deikman, Jill
 ; APPLICANT: Hardeman, Kristine J.
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Lalagudi, Raghunath V.
 ; APPLICANT: Ruan, Yijun G.
 ; APPLICANT: Ruft, Thomas G.
 ; APPLICANT: Sammons, R. Douglas
 ; APPLICANT: Shukla, Hridayabhiranjan
 ; APPLICANT: Wu, Kunsheng
 ; APPLICANT: Xu, Nanfei
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
 ; FILE REFERENCE: 38-21(51936)A
 ; CURRENT APPLICATION NUMBER: US/60/207,458
 ; CURRENT FILING DATE: 2000-05-30
 ; NUMBER OF SEQ ID NOS: 152403
 ; SEQ ID NO 33653

LENGTH: 1016
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; OTHER INFORMATION: Clone ID: uc-zmflb73263c09b1
 ; US-60-207-458-33653

Query Match 3.3%; Score 105.6; DB 64; Length 1016;
 Best Local Similarity 49.3%; Pred. No. 1.3e-05;
 Matches 452; Conservative 0; Mismatches 444; Indels 20; Gaps 6;

QY 1527 TATTAATTCGTTCTTAAATGCAATCTTAAACATTTATTTGTTAGTGAATTAATA 1586
 DB 79 TATTAATTCGTTCTTAAATGCAATCTTAAACATTTATTTGTTAGTGAATTAATA 138
 QY 1587 GAGCGAATCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1646
 DB 139 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 198
 QY 1647 CATACCTTGGATGTTAAACTTAATCTGTTCTCCAGGATTAATA-TTAATCAT 1704
 DB 199 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 258
 QY 1705 CGAGCTAAATAAATTTGCTTATTTTCGGAATGCAATGAAGATTAACCTTAATGACTT 1764
 DB 259 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 318
 QY 1765 AATTTTGAATAATGTAACCTTTTACATAGATTAATTAATTAATTAATTAATTAATTA 1824
 DB 319 TATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 377
 QY 1825 TAAATGACGCTC--TACAACTGTGATAGTCAATTTTTCGCAAAATTAATTAATGA 1881
 DB 378 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 437
 QY 1882 ATTGAATGCTACTGCTCAATGAATAAAGAGCTGATTAATTTAATTAAGACAA 1941
 DB 438 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 492
 QY 1942 ATTGTGAATAATGTAATTAATTTCTTACA-----ATATTAATAATAATGATGCTA 1993
 DB 493 ATATTTTAAATTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
 QY 1994 TAAATGATTCCTAATGCTTAATAAATTTTATTAATTAATTAATTAATTAATTAATTA 2053
 DB 553 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 612
 QY 2054 AACCAATTAATGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2113
 DB 613 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 672
 QY 2114 ATTAAATTTAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2173
 DB 673 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 732
 QY 2174 CTTCCAAATTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2233
 DB 733 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 792
 QY 2234 GAATGAAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2292
 DB 793 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 852
 QY 2293 CTCGAAGTATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2352
 DB 853 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 912
 QY 2353 CAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2412
 DB 913 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 972
 QY 2413 GTTTTCGACTATTTGA 2428
 DB 973 TTAATTAATTAATTAATTA 988

RESULT 13
US-60-245-228-17
; Sequence 17, Application US/60245228
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN CYCLASE PROTEINS, NUCLEIC
; ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
; FILE REFERENCE: C1000878
; CURRENT APPLICATION NUMBER: US/60/245,228
; NUMBER OF SEQ ID NOS: 2000-11-03
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 22350
; TYPE: DNA
; ORGANISM: HUMAN
US-60-245-228-17

Query Match
Best Local Similarity 45.2%; Score 97.6; DB 68; Length 22350;
Pred. No. 0.00027;
Matches 565; Conservative 0; Mismatches 669; Indels 16; Gaps 5;

1164 TTAGGTGAACAAATATGTATATGATGTCACATATACATATTCCTTACAGA 1223
1224 AAAAAAACAATTAAGAGAGTTACATATTCATATATGATGCTATACCTTCAGCTA 1283
4308 ATATCTAGTAT 4366
1284 TGCTATCTAGAGCTAAGAAATATATATATATATATATATATATATATATAT 1343
4367 AT 4426
1344 GTAAT 1403
4427 GTAAT 4486
1404 AAAAAAACAATTAAGAGAGTTACATATTCATATATATATATATATATATAT 1463
4487 AT 4546
1464 AAT 4606
4547 TAT 4663
1524 AAT 1583
4607 TAT 4663
1584 TAAAGAGCACTTGAACATATATATATATATATATATATATATATATATATAT 1643
4664 TAT 4723
1644 AT 1703
4724 AT 4783
1704 TCGAGTGAACAAATATATATATATATATATATATATATATATATATATAT 1763
4784 AT 4843
1764 TAAAT 1823
4844 AT 4902
1824 AT 1863
4903 TTAAT 4962
1884 TCAATGCTACTATCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943

4963 TAT 5022
1944 TTTTGAACAAAT 2000
5023 AT 5082
2001 TTTTGAACAAAT 2060
5083 AT 5142
2061 AAT 5202
5143 TAT 5262
2121 TTTTGAACAAAT 2180
5203 TAT 5262
2181 AT 2240
5263 AT 5314
2241 AGTAT 2300
5315 TAT 5374
2301 ATGAT 2360
5375 TTAAT 5434
2361 AAT 2410
5435 TTAAT 5484

RESULT 14
US-09-663-779-3024/C
; Sequence 3024, Application US/09663779
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Malvar, Thomas M.
; TITLE OF INVENTION: Shukla, Hridayabhijan
; ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; CURRENT APPLICATION NUMBER: 38-21(51376)P
; PRIOR FILING DATE: 05/09/663, 779
; PRIOR APPLICATION NUMBER: 60/154,678
; NUMBER OF SEQ ID NOS: 1999-09-17
; SEQ ID NO 3024
; LENGTH: 742
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-09-663-779-3024

Query Match
Best Local Similarity 48.38; Score 94.6; DB 26; Length 742;
Pred. No. 0.00046;
Matches 326; Conservative 0; Mismatches 344; Indels 5; Gaps 2;

1538 TTTTGAACAAAT 1597
713 TAT 654
1598 TAAAT 1657
653 TTTTGAACAAAT 594
1658 TTTTGAACAAAT 1713
593 TTTTGAACAAAT 534

Wed Jun 11 10:32:16 2003

us-09-502-426b-1_copy_1_3202.rmpm

Page 15

Db 66 NGG 64

Search completed: June 11, 2003, 05:00:41
Job time : 7696.77 secs

Query Match	100.0%	Score 3202:	DB 5;	Length 6888;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3202;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTGGTATTATATTGTGGTGGTGGTTGAGCTACAAATAATTAATTCGTGTTCTGGT	60	
Db	1	ATGTGGTATTATATTGTGGTGGTGGTTGAGCTACAAATAATTAATTCGTGTTCTGGT	60	
QY	61	TATTCGTTCACATGATTTGAGTTGGTCTCCAAATTGGATCCACGATAATTAATTT	120	
Db	61	TATTCGTTCACAGATTTGAGTTGGTCTCCAAATTGGATCCACGATAATTAATTT	120	
QY	121	AAATTCATTTAAATAATTTACAGTAATTAATATCTTCATTTGATTTGTTTAAACAA	180	
Db	121	AAATTCATTTAAATAATTTACAGTAATTAATATCTTCATTTGATTTGTTTAAACAA	180	
QY	181	AATATCTATCTTTGGTATATGAGAAAAATATGAGATTGGAATTTATATAATAAAGGAA	240	
Db	181	AATATCTATCTTTGGTATATGAGAAAAATATGAGATTGGAATTTATATAATAAAGGAA	240	
QY	241	TAATCGATTCATTGGTGGATTAACACAGTTTGTGTCTTTTGGTATATGT	300	

Db 241 TAATGATTCATTTGGTTGGATTACAGCTTAAGTTTTGTTGTTCTTTGTTATATGT 300
QY 301 AATAGTAATCAAAAGAGATATGATGAAGTGAACATATTTGTTAGACCCCA 360
Db 301 AATAGTAATCAAAAGAGATATGATGAAGTGAACATATTTGTTAGACCCCA 360
QY 361 AAAAAAAAAAACAACAACAAACCCCGGATATAGTTTTGTTGTTGAT 420
Db 361 AAAAAAAAAAACAACAACAAACCCCGGATATAGTTTTGTTGTTGAT 420
QY 421 AGTTAATTTGATCATATTAACATGATGATTTCTTGATTAATAGATTTGTTAC 480
Db 421 AGTTAATTTGATCATATTAACATGATGATTTCTTGATTAATAGATTTGTTAC 480
QY 481 CAATTAATTTGATCATATTTCTTGATTAATTAATTAATAGAGTGAATATCC 540
Db 481 CAATTAATTTGATCATATTTCTTGATTAATTAATTAATAGAGTGAATATCC 540
QY 541 GTTATATGATCTCATATCATATGATATGATTTCTGCTAATCCAGCAATTTATACA 600
Db 541 GTTATATGATCTCATATCATATGATATGATTTCTGCTAATCCAGCAATTTATACA 600
QY 601 AGAGTTGAGAAAAACGAAAAATAGAAAAAGGAGAGTGAACCATGAGATATG 660
Db 601 AGAGTTGAGAAAAACGAAAAATAGAAAAAGGAGAGTGAACCATGAGATATG 660
QY 661 TGAATTAATTAATCAAGAGATAGAGATGACAAACAAAGTTGGAATTAATGTCCT 720
Db 661 TGAATTAATTAATCAAGAGATAGAGATGACAAACAAAGTTGGAATTAATGTCCT 720
QY 721 GCCAGTTTCTCTCAGCAATCATATGACCCATTTTGGATTTTCTGATATTCGTTAAA 780
Db 721 GCCAGTTTCTCTCAGCAATCATATGACCCATTTTGGATTTTCTGATATTCGTTAAA 780
QY 781 TTTGCAATAGAGTTTGAATAATTTTATTTAGTGAATCTGAATTAATTTGTTCA 840
Db 781 TTTGCAATAGAGTTTGAATAATTTTATTTAGTGAATCTGAATTAATTTGTTCA 840
QY 841 GGTATTTGCAATATCTTCTGTTTAAAGCATATTTTCTTTTCTTTGTTCTCTCT 900
Db 841 GGTATTTGCAATATCTTCTGTTTAAAGCATATTTTCTTTTCTTTGTTCTCTCT 900
QY 901 TAACATATTAATTTCCGGATATATGATTAACAATGATATTCACAAACAAATGCTGG 960
Db 901 TAACATATTAATTTCCGGATATATGATTAACAATGATATTCACAAACAAATGCTGG 960
QY 961 ACCATTTGAATTAACATTTTCTCAACAATTCAGGACACTGACCTGCAATTAATA 1020
Db 961 ACCATTTGAATTAACATTTTCTCAACAATTCAGGACACTGACCTGCAATTAATA 1020
QY 1021 CGATTTTACAGGCTCACTAGTTGAGATTAAGTATAGCATTAAGCATTAAGACCCGTTCAAG 1080
Db 1021 CGATTTTACAGGCTCACTAGTTGAGATTAAGTATAGCATTAAGCATTAAGACCCGTTCAAG 1080
QY 1081 TATTTATACAAAGTTACAAAGTGAATATAGCTTGAATCCTTTAGAAAATTTTGAATTA 1140
Db 1081 TATTTATACAAAGTTACAAAGTGAATATAGCTTGAATCCTTTAGAAAATTTTGAATTA 1140
QY 1141 CCGGTTGTTATGTAATTAATAGATTTAGTGAACAAATATGTTATTCATTTAGTGCTCA 1200
Db 1141 CCGGTTGTTATGTAATTAATAGATTTAGTGAACAAATATGTTATTCATTTAGTGCTCA 1200
QY 1201 ACATATACATTAATTTCTTACAGAAAAACAACTTAAGAGAGTTAAACATATCCATATAT 1260
Db 1201 ACATATACATTAATTTCTTACAGAAAAACAACTTAAGAGAGTTAAACATATCCATATAT 1260
QY 1261 GGGTATGCTATACCTTTACAGTATGCTATAGAACTTAAGAAATAGTTATGATGTC 1320
Db 1261 GGGTATGCTATACCTTTACAGTATGCTATAGAACTTAAGAAATAGTTATGATGTC 1320
QY 1321 GATTAATGAATTTACAGCGGTGTAATTAATTAAGGACCGTATGTTAGATCACTGCAA 1380
Db 1321 GATTAATGAATTTACAGCGGTGTAATTAATTAAGGACCGTATGTTAGATCACTGCAA 1380

QY 1381 AATCATTTCTGTTGGTCAACAATAAAAAACAAGAAAAAGAAAGCATTTT 1440
Db 1381 AATCATTTCTGTTGGTCAACAATAAAAAACAAGAAAAAGAAAGCATTTT 1440
QY 1441 CTGGAATTCATTCATGATGATTAATAGATATCTTTGGTTACAGTTGGAAGTCC 1500
Db 1441 CTGGAATTCATTCATGATGATTAATAGATATCTTTGGTTACAGTTGGAAGTCC 1500
QY 1501 TCTACAGCGTGAACCATTCGACATTAATTAATGCTTTCTTAATGATCTTAACAT 1560
Db 1501 TCTACAGCGTGAACCATTCGACATTAATTAATGCTTTCTTAATGATCTTAACAT 1560
QY 1561 ATTATTTGTTAGTTGAATTTAATTAAGAGCAACTGTAACATTAATTAATTAAT 1620
Db 1561 ATTATTTGTTAGTTGAATTTAATTAAGAGCAACTGTAACATTAATTAATTAAT 1620
QY 1621 ATACTAGTATGATATTTCCAAATATCATCTTTGATGTTTAACTTAATCTGTTCT 1680
Db 1621 ATACTAGTATGATATTTCCAAATATCATCTTTGATGTTTAACTTAATCTGTTCT 1680
QY 1681 TCCTACGGTAAATTAATTAATCAATGAGTAAAAAAAGTTTGTCTTAATTTGCGCATG 1740
Db 1681 TCCTACGGTAAATTAATTAATCAATGAGTAAAAAAAGTTTGTCTTAATTTGCGCATG 1740
QY 1741 ATGAAGGATTAACCTTAATGATCTTAATTTTGAATAATGTAACCTTTTACTCATAGAT 1800
Db 1741 ATGAAGGATTAACCTTAATGATCTTAATTTTGAATAATGTAACCTTTTACTCATAGAT 1800
QY 1801 AATTACCGTATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1860
Db 1801 AATTACCGTATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1860
QY 1861 TCTGCAATTAATTAATTAAGAAATTCATGATCAATGAGAAAGAAACAGTGAAT 1920
Db 1861 TCTGCAATTAATTAATTAAGAAATTCATGATCAATGAGAAAGAAACAGTGAAT 1920
QY 1921 ACAATTTAATTTAAGCAAAATTTTGAATAATTTTGAATAATTTTGAATAATTT 1980
Db 1921 ACAATTTAATTTAAGCAAAATTTTGAATAATTTTGAATAATTTTGAATAATTT 1980
QY 1981 AATATGATGCTTAATGATTTTCTGATGTTCTTAATAATTTTCTTAATTTAGTTA 2040
Db 1981 AATATGATGCTTAATGATTTTCTGATGTTCTTAATAATTTTCTTAATTTAGTTA 2040
QY 2041 TAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2100
Db 2041 TAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2100
QY 2101 AAATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
Db 2101 AAATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
QY 2161 TTGTTAAGCAAACTTCCAAATTTTATTTTATGGAACAAATAATTAAGATGAGAAA 2220
Db 2161 TTGTTAAGCAAACTTCCAAATTTTATTTTATGGAACAAATAATTAAGATGAGAAA 2220
QY 2221 CTATTTTGTGTTGATGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
Db 2221 CTATTTTGTGTTGATGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
QY 2281 AGCCTATACGCGCTCAAGATATGTTATAGTATGATTAATTAATTAATTAATTAAT 2340
Db 2281 AGCCTATACGCGCTCAAGATATGTTATAGTATGATTAATTAATTAATTAATTAAT 2340
QY 2341 TCGAATTTGGGACAAACAAATGAGAAATTAATTAATTAATTAATTAATTAATTAAT 2400
Db 2341 TCGAATTTGGGACAAACAAATGAGAAATTAATTAATTAATTAATTAATTAATTAAT 2400
QY 2401 TTGAGTAAATGTTTCTGATATGAGGCGCAAAAAAGCAATGCAAAAGTCTAC 2460
Db 2401 TTGAGTAAATGTTTCTGATATGAGGCGCAAAAAAGCAATGCAAAAGTCTAC 2460

2461 GGGTTGACTGTCAGTGGGTAATATCTAATACTGCTTTGACCGCAGCTCGTG 2520
|||||
2461 GGGTTGACTGTCAGTGGGTAATATCTAATACTGCTTTGACCGCAGCTCGTG 2520
2521 TAGGGGCTCTGACATTTTACGCTGTTCTACCCCTACTCGTGAGCCACCTTTTCCCA 2580
|||||
2521 TAGGGGCTCTGACATTTTACGCTGTTCTACCCCTACTCGTGAGCCACCTTTTCCCA 2580
2581 TATCTTAAGGGTAATTTGGAAATCCCAATTTAAACCGATTTAGAGCCGAGCGGACTTC 2640
|||||
2581 TATCTTAAGGGTAATTTGGAAATCCCAATTTAAACCGATTTAGAGCCGAGCGGACTTC 2640
2641 TGGGATCTGCTGAGCATTTATCAAAAATTTATGACAGCAATGGGTTTATTAATTAA 2700
|||||
2641 TGGGATCTGCTGAGCATTTATCAAAAATTTATGACAGCAATGGGTTTATTAATTAA 2700
2701 AACCTACAATGATGATCAATTAATTTTCAAAACCTTTTACGATGATGCTGATGATCT 2760
|||||
2701 AACCTACAATGATGATCAATTAATTTTCAAAACCTTTTACGATGATGCTGATGATCT 2760
2761 ATCTAATGACTTTTCTTACACAGGCTGATGAAGTTATGATGATGATGATGATGATG 2820
|||||
2761 ATCTAATGACTTTTCTTACACAGGCTGATGAAGTTATGATGATGATGATGATGATG 2820
2821 ACAATTTGATATGATATATCCATTAATCCATGATATTTATGATATTAATAGCTGTTAA 2880
|||||
2821 ACAATTTGATATGATATATCCATTAATCCATGATATTTATGATATTAATAGCTGTTAA 2880
2881 CATATTGACATGCGAGCTTTCTGCACTTTTGTATTTAATTAAAGTTTAAATAATA 2940
|||||
2881 CATATTGACATGCGAGCTTTCTGCACTTTTGTATTTAATTAAAGTTTAAATAATA 2940
2941 AAGTATTAAGAGCATTAAGAGGCAACAAAGTAATGACAGGAGAGAGAGAGAGAGAG 3000
|||||
2941 AAGTATTAAGAGCATTAAGAGGCAACAAAGTAATGACAGGAGAGAGAGAGAGAGAG 3000
2941 AAGTATTAAGAGCATTAAGAGGCAACAAAGTAATGACAGGAGAGAGAGAGAGAGAG 3000
3001 TGAACCTCATGCTGATTTAGTTAAAGCTTAATAGAGATTTTATTAATTAAAGAGAG 3060
|||||
3001 TGAACCTCATGCTGATTTAGTTAAAGCTTAATAGAGATTTTATTAATTAAAGAGAG 3060
3061 ATAAATTTATTTTCTGACTTTCTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 3120
|||||
3061 ATAAATTTATTTTCTGACTTTCTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 3120
3121 TCAGTAGAAGTCCGATTCCTTAATTAAGCAAAAGCCATTAGAAAGAGAGAGAGAGAG 3180
|||||
3121 TCAGTAGAAGTCCGATTCCTTAATTAAGCAAAAGCCATTAGAAAGAGAGAGAGAGAG 3180
3181 GAG 3202
|||||
3181 GAG 3202

RESULT 2
US-09-502-426a-1
; Sequence 1, Application US/09502426A
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMP4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09502,426A
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA

; ORGANISM: Arabidopsis sp.
US-09-502-426a-1
Query Match 100.0%; Score 3202; DB 7; Length 6888;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGTGGGATTTATATGTTGGGTTGGTGGACCTACATTAATAATTTGGTTGGT 60
|||||
1 ATGTGGGATTTATATGTTGGGTTGGTGGACCTACATTAATAATTTGGTTGGT 60
61 TATTCGTTCACATGATTTGAGTTGGTTCATATTTGGATTTCCAGATTAATTAAT 120
|||||
61 TATTCGTTCACATGATTTGAGTTGGTTCATATTTGGATTTCCAGATTAATTAAT 120
61 TATTCGTTCACATGATTTGAGTTGGTTCATATTTGGATTTCCAGATTAATTAAT 120
121 AAAATTCATTTAAATTTTACAGTAATTAATTAATTTTACATGATTTATTAACA 180
|||||
121 AAAATTCATTTAAATTTTACAGTAATTAATTAATTTTACATGATTTATTAACA 180
181 AATATCTATCTTTGATATGAGAAATTAATGAGTTGGAAATTAATTAATTAAGGAA 240
|||||
181 AATATCTATCTTTGATATGAGAAATTAATGAGTTGGAAATTAATTAATTAAGGAA 240
241 TATCGATTCATTTGGTTGGATTTACACAGTTAAGTTTGGTTCTTTGTTATATGT 300
|||||
241 TATCGATTCATTTGGTTGGATTTACACAGTTAAGTTTGGTTCTTTGTTATATGT 300
241 TATCGATTCATTTGGTTGGATTTACACAGTTAAGTTTGGTTCTTTGTTATATGT 300
301 ATATGATTAATCAAAAAGATTTGATTTGAAGTGAACATATTTGTTATGACCCCA 360
|||||
301 ATATGATTAATCAAAAAGATTTGATTTGAAGTGAACATATTTGTTATGACCCCA 360
301 ATATGATTAATCAAAAAGATTTGATTTGAAGTGAACATATTTGTTATGACCCCA 360
361 AAAAAAAAAAAAAACAAACAAACACCCCGCCGATGATTTGGTTGGTCTGGAT 420
|||||
361 AAAAAAAAAAAAAACAAACAAACACCCCGCCGATGATTTGGTTGGTCTGGAT 420
421 AGGTTATTTGATCATTAATTAACATGATTTCTTTGATTTACTATGAGATTTCTTAC 480
|||||
421 AGGTTATTTGATCATTAATTAACATGATTTCTTTGATTTACTATGAGATTTCTTAC 480
481 CAATTAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 540
|||||
481 CAATTAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 540
481 CAATTAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 540
541 GTTATGATGATCTCCATCATGATTTGATTTGATTTGATTTGATTTGATTTGATTT 600
|||||
541 GTTATGATGATCTCCATCATGATTTGATTTGATTTGATTTGATTTGATTTGATTT 600
541 GTTATGATGATCTCCATCATGATTTGATTTGATTTGATTTGATTTGATTTGATTT 600
601 AGAGATTTGAGAAAAACCGAAATTAAGAAAAAGGAAAGAGTGAAGTGAAGTGAAG 660
|||||
601 AGAGATTTGAGAAAAACCGAAATTAAGAAAAAGGAAAGAGTGAAGTGAAGTGAAG 660
601 AGAGATTTGAGAAAAACCGAAATTAAGAAAAAGGAAAGAGTGAAGTGAAGTGAAG 660
661 TGAATTAATTAATCAAAAGATTAAGAGATGACAAACAAAGTTGGAATTAAGTCC 720
|||||
661 TGAATTAATTAATCAAAAGATTAAGAGATGACAAACAAAGTTGGAATTAAGTCC 720
661 TGAATTAATTAATCAAAAGATTAAGAGATGACAAACAAAGTTGGAATTAAGTCC 720
721 GCAGCTTCTGCAATCAATCAATATGACCCCTATTTGATTTCTGATATTTGATTTAA 780
|||||
721 GCAGCTTCTGCAATCAATCAATATGACCCCTATTTGATTTCTGATATTTGATTTAA 780
721 GCAGCTTCTGCAATCAATCAATATGACCCCTATTTGATTTCTGATATTTGATTTAA 780
781 TTTGCGATTAAGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 840
|||||
781 TTTGCGATTAAGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 840
781 TTTGCGATTAAGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 840
841 GGTATTTGATTAATCTTCTGTTTAAAGCATTTTCTTCTTCTTCTTCTTCTTCT 900
|||||
841 GGTATTTGATTAATCTTCTGTTTAAAGCATTTTCTTCTTCTTCTTCTTCTTCT 900
841 GGTATTTGATTAATCTTCTGTTTAAAGCATTTTCTTCTTCTTCTTCTTCTTCT 900
901 TACATTAATTAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 960
|||||
901 TACATTAATTAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 960
901 TACATTAATTAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 960
961 ACCATTTGAATTAATCTTTTCTCAAACTTAAGGAGAGTGAAGTGAAGTGAAGTGA 1020
|||||
961 ACCATTTGAATTAATCTTTTCTCAAACTTAAGGAGAGTGAAGTGAAGTGAAGTGA 1020

Dh 961 ACCATTTTGATTAACCTTTTCTCMAACATTAAGGACACTGGACCTCGACCTTAAATA 1020
Qy 1021 CGATTTTACAGCTCACTAGTGTAGATTAAGCATTAAGCATTAAGGACCCGTTCAAGC 1080
Dh 1021 CGATTTTACAGCTCACTAGTGTAGATTAAGCATTAAGCATTAAGGACCCGTTCAAGC 1080
Qy 1081 TATTTATACAAAGTACAAAGTAAATAGCTTGAATTCCTTGAAGAAATTTTGAATTA 1140
Dh 1081 TATTTATACAAAGTACAAAGTAAATAGCTTGAATTCCTTGAAGAAATTTTGAATTA 1140
Qy 1141 CGGTTGTTATGAATATATAGATTTAGTGTAAACAAATATGTAAATCAATTAAGTGTCA 1200
Dh 1141 CGGTTGTTATGAATATATAGATTTAGTGTAAACAAATATGTAAATCAATTAAGTGTCA 1200
Qy 1201 ACATATACATTAATTCCTTACAGAAAAACAACTTAAGAGAACTTAACATATCATATAT 1260
Dh 1201 ACATATACATTAATTCCTTACAGAAAAACAACTTAAGAGAACTTAACATATCATATAT 1260
Qy 1261 GGGTATGCTATACCTTTACAGTATGCTATATAGAGACTTAAGAAATATGTTATGATGTC 1320
Dh 1261 GGGTATGCTATACCTTTACAGTATGCTATATAGAGACTTAAGAAATATGTTATGATGTC 1320
Qy 1321 GATTAATGAATTCACAGCGGTGTATATATATGAGACGTAATGTTACATACATGCA 1380
Dh 1321 GATTAATGAATTCACAGCGGTGTATATATATGAGACGTAATGTTACATACATGCA 1380
Qy 1381 ATATCATTTCTTGGTGTCAACAAATTAACAAACAAACAAACAAACAAACAAACAAATTTT 1440
Dh 1381 ATATCATTTCTTGGTGTCAACAAATTAACAAACAAACAAACAAACAAACAAACAAATTTT 1440
Qy 1441 CTGGAATTCATTCATGATCTTAATGCTATAGATCTTTGGGTACAGTTTGAAGTCC 1500
Dh 1441 CTGGAATTCATTCATGATCTTAATGCTATAGATCTTTGGGTACAGTTTGAAGTCC 1500
Qy 1501 TCTACAGCGGTGTACACATCTGCACTATTAATGCTTCTTAATGCTTCTTAAT 1560
Dh 1501 TCTACAGCGGTGTACACATCTGCACTATTAATGCTTCTTCTTAATGCTTCTTAAT 1560
Qy 1561 ATTTATTTAGTGTGAATTAATTAAGAGCACTTGTAACTTAATTAATTAATTA 1620
Dh 1561 ATTTATTTAGTGTGAATTAATTAAGAGCACTTGTAACTTAATTAATTAATTAATTA 1620
Qy 1621 ATACTGATATGATATATTCCTTAATGCTTGTGATGTTTAACTTAATCTTGTCT 1680
Dh 1621 ATACTGATATGATATATTCCTTAATGCTTGTGATGTTTAACTTAATCTTGTCT 1680
Qy 1681 TCCATAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
Dh 1681 TCCATAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
Qy 1741 ATGAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
Dh 1741 ATGAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
Qy 1801 AATTAACGATATGTTTGTGCAATATGACAGCTTACAGCTGATATGATTAATTTT 1860
Dh 1801 AATTAACGATATGTTTGTGCAATATGACAGCTTACAGCTGATATGATTAATTTT 1860
Qy 1861 TCTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
Dh 1861 TCTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
Qy 1921 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
Dh 1921 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
Qy 1981 AATATATGCTAT 2040
Dh 1981 AATATATGCTAT 2040
Qy 2041 TAAAT 2100
Dh 2041 TAAAT 2100

Qy 2101 AAATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
Dh 2101 AAATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
Qy 2161 TTGTTAAACAGAACTTCAAAATTTTATTTTATGAGACAGAAATTAACAGATAGAAA 2220
Dh 2161 TTGTTAAACAGAACTTCAAAATTTTATTTTATGAGACAGAAATTAACAGATAGAAA 2220
Qy 2221 CTATTTGTGTGGAATGAGATATATATATATATATATATATATATATATATATATAT 2280
Dh 2221 CTATTTGTGTGGAATGAGATATATATATATATATATATATATATATATATATATAT 2280
Qy 2281 AGCCTATACGCGCTCAAAATATGTTATCTATGATGATTAATTAATTAATTAATTAATTA 2340
Dh 2281 AGCCTATACGCGCTCAAAATATGTTATCTATGATGATTAATTAATTAATTAATTAATTA 2340
Qy 2341 TCAGAAATGAGACAAATGAAACGGAATTAATTAATTAATTAATTAATTAATTAATTA 2400
Dh 2341 TCAGAAATGAGACAAATGAAACGGAATTAATTAATTAATTAATTAATTAATTAATTA 2400
Qy 2401 TTGAGTAAATGTTTCTGATATATGAGGGAACAAATGCAATGCAATGCAATGCAATG 2460
Dh 2401 TTGAGTAAATGTTTCTGATATATGAGGGAACAAATGCAATGCAATGCAATGCAATG 2460
Qy 2461 GGGTGTGACCTGCTCAGTTTCTGATATATATATATATATATATATATATATATATAT 2520
Dh 2461 GGGTGTGACCTGCTCAGTTTCTGATATATATATATATATATATATATATATATATAT 2520
Qy 2521 TAGGGTCTCTTGACACTTTTCACTGTTTCACTGTTTCACTGTTTCACTGTTTCACTG 2580
Dh 2521 TAGGGTCTCTTGACACTTTTCACTGTTTCACTGTTTCACTGTTTCACTGTTTCACTG 2580
Qy 2581 TATCTTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2640
Dh 2581 TATCTTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2640
Qy 2641 TGGGATTCGCTGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2700
Dh 2641 TGGGATTCGCTGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2700
Qy 2701 AACTCAACCTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2760
Dh 2701 AACTCAACCTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2760
Qy 2761 ATCTAATGACTTTTCTTCAACAGGATTAATTAATTAATTAATTAATTAATTAATTA 2820
Dh 2761 ATCTAATGACTTTTCTTCAACAGGATTAATTAATTAATTAATTAATTAATTAATTA 2820
Qy 2821 ACAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
Dh 2821 ACAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
Qy 2881 CTATTTGAGATGCGAGCTTCTGCACTTTTGTATTAATTAATTAATTAATTAATTA 2940
Dh 2881 CTATTTGAGATGCGAGCTTCTGCACTTTTGTATTAATTAATTAATTAATTAATTAATTA 2940
Qy 2941 AAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3000
Dh 2941 AAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3000
Qy 3001 TGAAGCTCATGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
Dh 3001 TGAAGCTCATGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
Qy 3061 ATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120
Dh 3061 ATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120
Qy 3121 TCAGTAAAGTCCGATTTCCCATTTTAAGACAAAGCCATTTAGAAAGAGAAAGTGA 3180
Dh 3121 TCAGTAAAGTCCGATTTCCCATTTTAAGACAAAGCCATTTAGAAAGAGAAAGTGA 3180

RESULT 4
US-10-312-841-1
: Sequence 1, Application US/10312841
: GENERAL INFORMATION:
: APPLICANT: Epigenomics AG
: TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
: FILE REFERENCE: E01/1208/WO
: CURRENT FILING DATE: 2002-12-30
: NUMBER OF SEQ ID NOS: 2
: SEQ ID NO 1
: LENGTH: 3673778
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
: NAME/KEY: unsure
: LOCATION: (3294164)
: US-10-312-841-1

3.2%: Score 103; DB 9; Length 3673778;

Best Local Similarity 42.4%; Pred. No. 6.1e-07;
Matches 989; Conservative 0; Mismatches 1320; Indels 24; Gaps 7;

QY 1 ATGTGGTATTAATTTGGTGGTTCGCTTACACATTAATTAATTCGTTCTGCT 60
DB 1713239 ATTTATTTTATATATATATATATATATATATATATATATATATATATG 1713298
QY 61 TATTCGTACATGATTTAGTTGGTCTCATTTGGATTCACATATATTAATTT 120
DB 1713299 TTTTATATTTTATATATATATATATATATATATATATATATATATTTT 1713358
QY 121 AAAATCATTTAAATTTTCAAGATTAATTAATTAATTCATTTGATTTGATTAACA 180
DB 1713359 TATATATATTTTATATATATATATATATATATATATATATATATATATAT ---TT 1713415
QY 181 AATATCTATCTTGGTATATGAGAAATATGAGTTGGAATTTAATTAATAAGAAA 240
DB 1713416 TATATATATTTTATATATATATATATATATATATATATATATATATAT 1713475
QY 241 TAATCGATTCATTTGGTGGATTAACAGTTAAGTTTGTGTTCTGTTATATG 300
DB 1713476 TTTATATATATTTTATATATATATATATATATATATATATATATATAT 1713535
QY 301 ATATGATTAATCAAAAAGATTTGATGAGTGAACATTAATTCGTTATGACCCCA 360
DB 1713536 ATTTTATATATTTTATATATATATATATATATATATATATATATATAT 1713595
QY 361 AAAAAAAGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 420
DB 1713596 ATTTATTTTAT 1713655
QY 421 AGGTTATTTGATCATTAATACATGATCATTTCTTTGATTACATGAATTTCTTAC 480
DB 1713656 TTTATATATATTTTATATATATATATATATATATATATATATATATATATA 1713715
QY 481 CAATTAATTTGCAATTCATCTCTGATTTAATTAATTAATACAGTGAATATCC 540
DB 1713716 TA 1713775
QY 541 GTTTATCGATCTCCATCATGATTAATGATCTTGTGCTAATCCAGCAATTAATACA 600
DB 1713776 TATTTATATATATTTTATATATATATATATATATATATATATATATAT 1713835
QY 601 AGAGTTATGAGAAAAACGAAAAATAGAAAAAGGAAGAGTGTGACCATGGAGATG 660
DB 1713836 TATATATTTTAT 1713895
QY 661 TGAATATTAATCAAGAGATA-AGAGATGACAACAAAAGTTGGAATATATGCTCC 719
DB 1713896 TTTATATATATTTTATATATATATATATATATATATATATATATATATATAT 1713955

QY 720 TGCCAGCTTCTCTCACATCAATATGACCCATTTTGGATTTTCTGGAAATTCGTTAAA 779
DB 1713956 ATATATATTTTAT 1714015
QY 780 ATTTGGCATATACGATTTGCAAAAAATATTTATTTGTTAGCTGATCTCAATATATGTTCC 839
DB 1714016 TATATATTTTAT 1714075
QY 840 AGGATTTTGCATATCTCTGTTTAAAGCATATTTGCTTTTGGTTGCTTCTC 899
DB 1714076 ATATATTTTA 1714135
QY 900 TTAACATATATATATGCGGATATATGATTAACAATGATATATCAAAACAAATTCGTCG 959
DB 1714136 TATATTTTATATA-----TATATTTTATATATATATATATATATATATATAT 1714189
QY 960 GACATTTTGAATTAACCTTTTTCGCAACATTTACGGGACCTGACCTGACCTTAAT 1019
DB 1714190 ATATATATTTTATATATATATATATATATATATATATATATATATATAT 1714249
QY 1020 ACGATTTTACAGCTCATGATTTGAGATTTACTATACATA---AAGCATTAAGACCCGCTC 1076
DB 1714250 TTTTATATATATTTTATATATATATATATATATATATATATATATATATATAT 1714309
QY 1077 AAGCTATTTATACAAAGTTACAAACGTAATATAGCTTGAATCCTTTAGAAAATTTTGA 1136
DB 1714310 TATATATTTTAT 1714369
QY 1137 ATTAACGGTGTATATGTAATATATATATATATATATATATATATATATATATAT 1196
DB 1714370 ATATATATTTTAT 1714429
QY 1197 GTCACAT 1256
DB 1714430 TATATATTTTAT 1714489
QY 1257 ATATGGTAT 1316
DB 1714490 ATATATTTTAT 1714549
QY 1317 TGTGCAATTAATGAAATTCACACGCGGTGATATATATATGAGACCTGTTAGCATC 1376
DB 1714550 TATATATTTTAT 1714602
QY 1377 GCAAAATATCATCTTGTGTTGTCGCAACATTAATAAAGAAAAAGAAAAAGAAAT 1436
DB 1714603 ATATATATATTTTATATATATATATATATATATATATATATATATATATAT 1714662
QY 1437 TTTTCTTGGAATTCATTCATATGATCTTAATATGATATGATCTTTGGTTACAGTTG 1496
DB 1714663 ATATTTTAT 1714722
QY 1497 GTCCTTACAAAGCGTGAACATCTGCACATTAATTAATTTCTTTATATGATCTTTA 1556
DB 1714723 TGTATATTTTAT 1714782
QY 1557 ACATATTTATTTAGTTTGAATTTAATAAGAGCAAGCTTGAACATTAATTAATTA 1616
DB 1714783 TTTATATATTTTAT 1714842
QY 1617 TTGATATCTAGTATGATATATTTCCAAATATGATCTTTGGATGTTTAACTTAATCT 1676
DB 1714843 TATATATTTTAT 1714902
QY 1677 TTTCTTCCAGGATTAATATTTATATATATATATATATATATATATATATATAT 1736
DB 1714903 TTTTATATATATTTTATATATATATATATATATATATATATATATATATAT 1714962
QY 1737 ATGCATGAGGATTAACCTTAATGATCTTAATTTTGGAAAA---TGTAACTTTTATCTC 1793
DB 1714963 ATATATATTTTAT 1715022

QY	1794	ATGATTAATTAACCGTATGTTTTGGTCCATATAGACAGCCTCCTACACCTGATGTC	1853
Db	1715023	ATATATATATTTATATATATTTTATATATATATATATATATATATATATATTTTAT	1715082
QY	1854	AAATTTTTCGCCAAATATTAATTTAGGAATTCAGTCTACTATCAATAGAGAAACAGCT	1913
Db	1715083	ATATATATATTTTATATATATATATTTTATATATATATATATATATATATATATAT	1715142
QY	1914	GAGATATACATTTAATTTTAAAGCAAAATTTTGAAGAAATGTTATATTTCTAACATA	1973
Db	1715143	ATATTTTATATATATTTTATATATATATATATATATATATATATATATATATATAT	1715202
QY	1974	TTATTTAAATATGATGCCATTAATGTTATTCATGTTCTTAAATATTTTTTTATAT	2033
Db	1715203	TATATATATATTTTATATATATATTTTATATATATTTTATATATTTTATATATTTATATA	1715262
QY	2034	TTAGTTTAAATNCATATATGACCAATAATATAGTTGGGAATCCAAATCTCCATTATA	2093
Db	1715263	TATATTTTATATATATTTTATATATATATATTTTATATATATATTTTATATATATTTT	1715322
QY	2094	TTTTTUGAAATCTCAAAATTTATTAATTTTATGTCATTAACCAATCATAGAAAGTCCAAA	2153
Db	1715323	ATATATATATTTTATATATATATTTTATATGATATTTTATATATATATATATTTTATAT	1715382
QY	2154	AAAAATTTTGTTAACAGAACTCCAAATTTTTTTTTTATGACACAGAAATATACAGA	2213
Db	1715383	GTAATTTTATATATATATTTTATATATGTTATTTTATATGATATATTTTATATATATTTT	1715442
QY	2214	TAGAAACATATTTTGTGTGCGAATGGAAGTATATATTTTATCTATTAAGCAAAATTTTAAAAA	2273
Db	1715443	TATATGATATTTTATATATATATTTTATATATGCTATATATTTTATATATATATATATATA	1715501
QY	2274	TTATATTAAGCCTATACGCCCTCAAGATATGTATCTCTAGTGGCTAATTTATA	2326
Db	1715502	TTTTATATATATATATTTTATATGCTTTGTTTTTTTGGAAAAATTTGGATTGATA	1715554

```

RESULT 5
US-10-312-841-1/c
: Sequence 1, Application US/10312841
: GENERAL INFORMATION:
:   APPLICANT: Epigenomics AG
:   TITLE OR INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
:   FILE REFERENCE: E01/1208/WO
:   CURRENT APPLICATION NUMBER: US/10/312,841
:   CURRENT FILING DATE: 2002-12-30
:   NUMBER OF SEQ ID NOS: 2
:   SEQ ID NO 1
:     LENGTH: 3673778
:     TYPE: DNA
:     ORGANISM: Artificial Sequence
:     FEATURE:
:     OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
:     FEATURE:
:     NAME/KEY: unsure
:     LOCATION: (3294164)
:     US-10-312-841-1

```

Query Match	3.1%;	Score 100.2;	DB 9;	Length 3673778;
Best Local Similarity	43.4%;	Pred. No. 1.6e-06;		
Matches 946;	Conservative	0;	Mismatches 1218;	Indels 15; Gaps 10;

OY	111	TTTTAAATATTTAAATTCATTTTAAAAATTTTACAGTAATATATATATCTTCACAT-TGTAT	169
Db	1715393	ATTTAAATATTCATATTTAAAAATATATATATTTAAAAATATTCACATATTTAAAAATATATATATTTAA	17153344
OY	170	TGTTATTAACAAAAATATCTATCTTGGATATATGAGAAATATATGAGTTTGGATTTTAAAT	229
Db	1715333	AATTTATATATATATAATATATATATATAAATATATATTTAAAAATATATATATATAAATATATATA	17152744
OY	230	AATTAAGAAATATATCATTCATCTTTGGTGGATTACACAGTAAAGTTTTGCTTCTT	289
Db	1715273	TATATAATATAT-T-ATATATAAATATATATATATAAATATATATATATAAATATATATATATAA	17152155

OY		290	TTCGTTATGTCTAATGCTAAACAAAAGATTCATTGAAGTGTAACAATATTCGT	349
Db		1715214	TATATATATATATAAATATATATATAAAATATATATATATAAAATATATATAAATATA	1715155
OY		350	TATGACCCCCAAAAAAAAAAAAAAACAAACAACAAACCCCCCGATATAGTTTTT	409
Db		1715154	TATATATAAATATATATATATAAATATATATATATAAATATATATATAAATATATA	1715095
OY		410	GGTTCGCGATTAGTGTATTTTGATCATTAATTACATGCATCATTTCTTGATTACTATGAA	469
Db		1715094	AAATATATATATATATAAATATATATATATAAATATATATATATAAATATATATATA	1715035
OY		470	GATTTCTCACAATTAATAATTTGCAATTCATATCTTGATTAATTAATAATACGAG	529
Db		1715034	AAATATATATATATATAAATATATATAAATATATATATATAAATATATAAATA-TAT	1714976
OY		530	TGTCATATCCGTTTATTCATCACTCCAATCATGATTATGATTCCTTGCTAAATCCAGCA	589
Db		1714975	ATATAAATATATATATAAATATATATATATACAAAATATATATATACAAAATATATATA	1714916
OY		590	AATATATACAGAAGTATTCGAGAAAAACGGAAATTAAGAAAAAGGGAAGAGTAGCAC	649
Db		1714915	AAATATATATAAATATATATATAAATATATATATAAATATATATAAATATATATATACA	1714856
OY		650	CATGAGATATGTGAATTAATATCAAAAGATATAGAGATGACAACAACAAAGGTTCGAA	709
Db		1714855	TATATAATATATATATATATATATATAAATATAAATATAATATATATAAATATATATAT	1714796
OY		710	TAAATGCCCTCCAGCTTCTCTCAACATCATATGACCCATTTGGATTTCGTGAT	769
Db		1714795	AAATATATATATAAATATATATATAT-AAAATATACATATATAAATATATATATAAATATAC	1714737
OY		770	ATTCGTTAAATTTGCGATPAACGATTCGAAAATATTTATTTGTAGCTGATCTCAT	829
Db		1714736	ATATAAATATATACATATAAATATATACATATAAATATACATATAAATATACATATAA	1714677
OY		830	ATTATGTTCCAGTATTTGCAATATCTTCTGTTTAAAGCATNTTGTCTTTCTTTTGT	889
Db		1714676	TATATATAAATATATATATAAATATATATATATAAATATATATATAAATATATATA	1714617
OY		890	TTCGTTCTGTACATATATATATTCGCGGATATATGATPAACATGATATATCACAAC	949
Db		1714616	TAAATATATATATATAAATATATACATATAAATATATATATAAATATATATACATATA	1714557
OY		950	AATGTCGTGGACCAATTTGAAATPAACTTTTCTCAACAATACGAGCACTGCATCGA	1009
Db		1714556	ATATATATATAAATATATATATAAATATATATAAATATATATAAATATATATATA	1714497
OY		1010	CCCTTAATAATACGATTTTACAGCTCACTAGTTAGATTTACTAGCATATAAGCATATAAGGA	1069
Db		1714496	AAATATATATAAATATATATAAATATATATAAATATATATAAATATATATATATA	1714437
OY		1070	CCCCTTCAGCTATTTATACAAAGTTPCAAAGCTGAATATAGCTTGAATCCTTTAGAGAA	1129
Db		1714436	ATATATATAAATATATATAAATATATATATAAATATATATAAATATATATATAA	1714377
OY		1130	TTTTGGAATTACCGCGTTGTTATGT-AAATATAGATTAGTGGTAACAACAATATGTATATC	1188
Db		1714376	TATATATATATAAATATATATAAATATATATAAATATATATAAATATATATATA	1714317
OY		1189	AATATAGTGGCAACATATACATAATTCCTTACAGAAAAACAACCTTAPAGGAAGTAAAC	1248
Db		1714316	ATATATATAAATATATAAATATATATATAAATATATATATAAATATATATAAATATA	1714257
OY		1249	ATATCCATATATGSGTATGCTATACCTTTCAACGTATGCTATA-CTAGAGACTPAAGATA	1307
Db		1714256	TATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATATATA	1714197
OY		1308	GTTATGTGATGCGATATAAGAAATTCACGCGGTATATATTTATGGAGCCGTATGCT	1367
Db		1714196	TATATATAAATATATATAAATATATATAAATATATATATAAATATATATATAAATATAT	1714137

OY 1368 ACGATCATCGAAAAATCATCTTGGTGGTCACAAATAAACAAAACAAAGAAAAA 1427
 Db 1714136 ATATATAATATATATATAATATATATATAATATAATATAATATAATATAATATA 1714077
 OY 1428 GAAAAAGATTTTCTTGCGATTCATTCATGATGCTCAATGCAATGATCTTTGGGTAC 1487
 Db 1714076 TATATAATATATATATAATATAATATAATATAATATAATATAATATAATATA 1714018
 OY 1488 AGTTTCGAGTCTCTACAGCGGTAAACATCTGCAACTATTAATGCTTCTTTAAT 1547
 Db 1714017 TATATAATATATATATAATATAATATAATATAATATAATATAATATAAT 1713958
 OY 1548 GCATCTTAAATATTTTGTGTAGTTGAGTTTAAATAGACGCACTTGTACATTCA 1607
 Db 1713957 ATATATAATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713898
 OY 1608 ATATTATATATAGATA-CTAGTATGTGATTTATTCCAATACATCTTTGAGTTTAAC 1666
 Db 1713897 AAATATATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713838
 OY 1667 TTAATCTTGTCTCTCAGCGTAAATATTAATCATCGAGTAAAAAAGTTTGTCT 1726
 Db 1713837 TAAATATAATATATATAATATAATATAATATAATATAATATAATATAATATA 1713778
 OY 1727 TATTTCGCGATCGATGAAGATTAACCTATAGCTTAATTTTGTGAATGTAAACC 1786
 Db 1713777 TATATAATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713718
 OY 1787 TTATCATCATGATTAATTCAGCTATGTTTGTGTCATTAATGACAGCCTCTACACGT 1846
 Db 1713717 TATATATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713661
 OY 1847 GATAGCTCAATTTTTCGCAAAATATTAATAGAGATTCATGCTACTATCAATGAGA 1906
 Db 1713660 TATATAATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713601
 OY 1907 AACGCGTATGATTAATTAATTAAGCAAAATTTTGAATAATGTTATATTTCT 1966
 Db 1713600 TATATAATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713541
 OY 1967 AACATATATTAATAATATGATGCTATTAATGATATTCCTATGTTCTTAATAATTTT 2026
 Db 1713540 AAAATATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713481
 OY 2027 TTATATTTAGTATAAATACATTAATGCAATTAATAGTTGGTAATTCAAATATCTCC 2086
 Db 1713480 TATATAATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713421
 OY 2087 ATTAATATTTTGTGAATCTCAAAATATTAATTTTGTGCAATTAACATGATAGAAAG 2146
 Db 1713420 ATATAATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713361
 OY 2147 T-----TCCAAAAAATTTTGTAAACAGAACTTCCAAATTTTATGAGACAA 2202
 Db 1713360 TAAATATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713301
 OY 2203 GAAATTAACAGATAGAAAACTATTTTGTGTGGAATGGAAGTAGTAATATACATTAAGCA 2262
 Db 1713300 AACATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 1713241
 OY 2263 ATTTTAAAAAATTTATATA 2281
 Db 1713240 ATATAATAATATAATATA 1713222

RESULT 6
 US-09-948-124-61/C
 : Sequence 61, Application US/09948124
 : GENERAL INFORMATION:
 : APPLICANT: VENTER, J. Craig
 : TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON CH
 : FILE REFERENCE: C1001301
 : TITLE OF INVENTION: Y, METHODS OF DETECTION, AND USES THEREOF

```

: CURRENT APPLICATION NUMBER: US/09/948,124
: CURRENT FILING DATE: 2002-12-23
: NUMBER OF SEQ ID NOS: 183
: SEQ ID NO 61
: LENGTH: 318007
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)_(318007)
: OTHER INFORMATION: n = A,T,C or G
US-09-948-124-61

```



```

OY 1634 ATTTATCCAAATACACACTCTTGATGTTTAAACCTAACTCTGTTCTCCACGGTAA- 1692
Db 500660 TTATTTTATATACCTATTTATACATAAATAAAATATATTTATTTATATATATAC 500601
OY 1693 -AATATTATATCATCGAGGTAAAAAAGTTTGTCTTATTTTCGGAGTACATGAAGATPA 1751
Db 500600 TAAATTTTATATATTTATTAATATATATTTATATATATTTATATATATATATAT 500541
OY 1752 ACCATATACCTTATTTTGGAAATGTAACCCCTTTACATAGATTAATACCGTAT 1811
Db 500540 AATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 500481
OY 1812 GTTTTGTTCGCCAATATGACAGCCCTCTACACTGTAAGTACATTTTTCGAAATAT 1871
Db 500480 AAATATATATGTTATTTATATATATATATATATATATATATATATATATATATAT 500421
OY 1872 -----TAAATTAGAAATTCATGCTACTATCATAGAAAGAAACAGCTGAGTATTACATT 1926
Db 500420 ATTTATTAATATATATTTATTAATATATATATATATATTTATTTATTAATATATATAT 500361
OY 1927 TAAATTTAAGCAAAATTTTGGAAAAATGTTATATTTCTTACAAATATTAATAAATG 1986
Db 500360 ATATATTTTATATATATTTTATTAATATATATATTTTATATATATTTTATATAT 500301
OY 1987 ATGCCATTAATGTAATTTCCATGTTCTTAAATATTTTTTTATTTTACTTTAATA 2046
Db 500300 TATATATTTATTTATTAAGTATATATATTTATTTATATATATTTATTTAATATATA 500241
OY 2047 CATATATGACCAATATAGTTGTCGAATTCAAATATCCATTAATTTTGTAAATCT 2106
Db 500240 TA-ATATATGCAATATATATATATATATATTTTAAATATTTTATATATATATATAT 500182
OY 2107 ACAATTTATTAATATTTAGTCATATACATGATAGAAAGTTCAGAAAAAATTTTGTTA 2166
Db 500181 ATTTATTTTATACATATATGATATTTATTTATTAATAATATGATATTTTATTTATCATATAG 500122
OY 2167 ACAGAACTCCAAATTTTTTTTTTATATGACAGAAATATACAGATAGAAACTATTT 2226
Db 500121 ATATTTCTTATTAATATATATATGTTCTTCATATGATATGATATTTTATTAATAATATAT 500062
OY 2227 TGTGTGCAATGAGATAGTATATATACATTAAGCAAAATTTTAAAAAATATATATAGCCTA 2286
Db 500061 TTCTTTATTAATATATTTATATATATATATATATATATATATATATATATATATAT 500002
OY 2287 TACGGCTCAAGTATGTTATCTAGTA 2313
Db 500001 TATATATTAATATTAATATATATATA 499975

RESULT 8
US-09-947-911-331
; Sequence 331, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 331
; LENGTH: 2453797
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2453797)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-331

Query Match 3.08; Score 97.4; DB 7; Length 2453797;
Best Local Similarity 44.98; Pred. No. 4.3e-06;
Matches 469; Conservative 0; Mismatches 361; Indels 15; Gaps 2;

```


Db	40582	ATAATATATATACCTTAATAATATATATATCATATATGAATATATATTAATATAGATCTA	40641
Oy	2125	GTCAAATACAATGATAGAAAGTCCAAAAAAAATTTGTACACAAAATCCCAATT	2184
Db	40662	TATCAAAATATATATATATTAATATATATATATATATTAATTTAAATATATATAGATATATAC	40701
Oy	2185	TTTTTTTTTATGCAACAGAATAATACAGATAGAAAACATATTTTGTTGGTAATGGAAGTA	2244
Db	40702	TATATGAAATATACCTTTATATATATGATATATACATATATGAAATATAAATTTTAAATATATG	40761
Oy	2245	GTATATATACATTAAGCAAATTTT-----AAAAATTTATATAGCCTTACGCCGTCAAAAT	2300
Db	40762	AATATATATATATATATGATATATATGACAAACATATATATGATATATATATATATATACGAAT	40821
Oy	2301	ATGTTATCTAGTAGCGTGTAATTAATAATATGCATGTCGATTCGAATTTGGGACMACAATG	2360
Db	40822	ATATATAATATATAGATATATATGAATATATATTAATAATATATATATATATATATATATG	40880
Oy	2361	AAAACGAATTTAAATATTTAACCTTTAAATTAATTAATTAATTAATTTGAGTAAT	2410
Db	40881	AATATATTAATATAGATATATATATATATGAAATATATTAATATATATATATATATAT	40930

```

RESULT 11
US-09-948-124-61
: Sequence 61, Application US/09948124
:
: GENERAL INFORMATION:
:
: APPLICANT: VENTER, J. Craig
:
: TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON CH
:
: TITLE OF INVENTION: Y, METHODS OF DETECTION, AND USES THEREOF
:
: FILE REFERENCE: CLO01301
:
: CURRENT APPLICATION NUMBER: US/09/948,124
:
: CURRENT FILING DATE: 2002-12-23
:
: NUMBER OF SEQ ID NOS: 183
:
: SEQ ID NO 61
:
: LENGTH: 318007
:
: TYPE: DNA
:
: ORGANISM: HUMAN
:
: FEATURE:
:
: NAME/KEY: misc-feature
:
: LOCATION: (1)...(318007)
:
: OTHER INFORMATION: n - A,T,C or G
:
: US-09-948-124-61

```

Query Match	3.0%;	Score 94.8;	DB 6;	length 318007;
Best Local Similarity	46.5%;	Pred. No. 1.4e-05;		
Matches 411;	Conservative	0;	Mismatches 467;	Indels 6; Gaps 3

QY	1530	TAAATGCTTCTTAAATGCACTTTAAACATATTTATGTTACTGGAAATTTATATAGAC	1589
Db	133360	TATATATATATATATATATTTAAATTTTAAATATATATATATATATATATATATATAT	133419
QY	1590	CGAAGCTGTAAACATACAAATATTTATATAGATACAGATGAGATTTATCCAAATACAT	1649
Db	133420	ATGATATATTAATATATTAATATATTTATTTAAAGAAATATATATTTATATAAATATCTAT	133479
QY	1650	ACTTTGATGTTTAAACCTTAATCTTGTCTTCCTCGAGGTATTAATATTAATCATCGAGG	1709
Db	133480	ATGATGTAAGAAACATATATTTATATAGTAAATATCTATATGATATAAATAAATATACAT	133539
QY	1710	TAAAAAAAGTTTGTCTTATTTTCCGAGATCGTAAGAGATTAACCTTAATGACTTTAATTT	1769
Db	133540	ATTTTATTAATTAATATCTATATAGTATTAATTAATATATATATTTAATATATTAATAATAT	133599
QY	1770	TTTGAATAATGTAACCTTTTACTACATAGATTAATTAACGATATGTTTGTGGCATTAATG	1829
Db	133600	TTAAATATATTTATATCATATATATCTATATATTTATATATATTTATTAATTAATATATAT	133659
QY	1830	ACAGCCTCTACAACTGTGATGTCATATTTTCTCGAAATATTAATTAATGAGATTCAGAG	1889
Db	133660	ATTAATTAATATATATCTTATTAATTAATATTAATTAATTAATATATATTAATAAATATATATAA	133719
QY	1880	CTACTATCATATAGAAAGAACGCTGAGATTTTACATTTTATTTAAAGCAAAATTTTGA	1949

Dd	133720	TAAATATATTTTAAAAAATATATA - AATATATTTTTTAATAATATGATATTTTAAA	133777
Oy	1950	AAAAATGTATAATTTCTAACAAATATTTAAAAATAGACCTATATGTTTCCTAG	2009
Dd	133778	TAAATATATATTTTATTAATATATATTTAATATATATGATGAATAAATACATAT	133837
Oy	2010	TTCTTA - AAATTTTTTTTTTATATTTAGTATAAATACAT --- TATGAACAATAATAG	2065
Dd	133838	TATTTATATATATTTTATTAATAACATAATATATTTAATATATATATTTTTTTAT	133897
Oy	2066	TTGTGAATTCAAATATCCATTAATATTTTTTGAAATCTCAAAATATTAATATTAG	2125
Dd	133898	AATATATATTTTATATATTTTTTATATATATTAATATATATATATATATATATTA	133957
Oy	2126	TCAATACAAATGCATAGAAGTTCMAAAAAAAAAATTTGTTAACAGAACTTCCAAATTT	2185
Dd	133958	AATATATTAATATATTTTATTAATATATATAAATTTGTAATATTAATAAATTAATATA	134017
Oy	2186	TTTTTTTTTATGSAACAGAAATACAGATAGAAAATATTTGTTGTGGAAATGGAAGTAG	2245
Dd	134018	TTTTTATTTATGTATAAATATACATATATATAAATAATTAATACAAATTAATATATAATA	134077
Oy	2246	TAAATACATTAAGCAAAATTTTAAAAAATTAATATAGCCTATACGGCTCAAAATATGTT	2305
Dd	134078	TAAATATATATATTAATATATATATATCATATTTTAAATATATTAATATATTTATTTATAG	134137
Oy	2306	ATCTAGTAGTGCTAATTAATAATAGCATGGCGATTCAGAAATGGGACAACAATGAAAAC	2365
Dd	134138	AATTAATATCAATATATATATATATCTGAATTAATATTAATAACAATATATTTACATAT	134197
Oy	2366	GGAATTAATAATTTTACCTTTAAATAAATAAATAATTTGGCTAAA	2409
Dd	134198	AATATAAATATGCTAATTAATATATATATCATGTATATATATATAA	134241

```

RESULT 12
US-09-947-911-2
: Sequence 2, Application US/09947911
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig
: TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED
: TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
: FILE REFERENCE: CL001300
: CURRENT APPLICATION NUMBER: US/09/947,911
: CURRENT FILING DATE: 2001-09-07
: NUMBER OF SEQ ID NOS: 368
: SEQ ID NO 2
: LENGTH: 600322
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(600322)
: OTHER INFORMATION: n = A,T,C or G
US-09-947-911-2

```

Query Match	3.0%;	Score 94.8;	DB 7;	Length 600322;
Best Local Similarity	46.5%;	Pred. No. 1.3e-05;		
Matches 411;	Conservative	0;	Mismatches 467;	Indels 6; Gaps 3;

QY	1530	TAAATGCTTCTTTAAATGCACTCTTACATATTTATCTGACTGGAAATTTAAATAGAC	1589
QY	1530	TAAATGCTTCTTTAAATGCACTCTTACATATTTATCTGACTGGAAATTTAAATAGAC	1589
Db	499968	TATATTATATATATATTTAAATATTTAAATATATATATATATATATATATATATATAT	500027
QY	1590	CGAAGCTTGACATACATATTTATATATAGATACAGATATGATTTATCCAAATACAT	1649
Db	500028	ATGAT	500087
QY	1650	ACTTGAGGATTTAACTTAATCTGTTCTCCACGCTATATATATATATATATATATATAT	1709
Db	500088	ATGATATGAGAACTATATATTTATATAGTAAATATCTATATGTTATATATATATATATAT	500147

Query Match	2.98;	Score 94.2;	DB 7;	Length 597573;
Best Local Similarity	45.28;	Pred. No. 1.6e-05;		
Matches 519; Conservative	0;	Mismatches 608;	Indels 20;	Gaps 4;

[illegible]

QY	2294	AGTATATTACACTTTAAGCAAACTTTTAATAAAAATATATATAACCGCATACGCCGTCAAATATG	2303
Dd	238091	ATTATATATATATATGCTTAATATTCATTATATATATATATAATATATACATATATATATATTATA	238092
QY	2304	TTCATCCTA	2310
Dd	238091	TAAATATA	238095

```

RESULT 14
US-09-947-911-323/C
: Sequence 323, Application US/09947911
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig
: TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON CH
: TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
: FILE REFERENCE: C1,001,300
: CURRENT APPLICATION NUMBER: US/09/947,911
: CURRENT FILING DATE: 2001-09-07
: NUMBER OF SEQ ID NOS: 368
: SEQ ID NO 323
: LENGTH: 4063095
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(4063095)
: OTHER INFORMATION: n = A,T,C OR G
: US-09-947-911-323

```

Query Match	2.9%;	Score 92.4;	DB 7;	Length 4063095;
Best Local Similarity	24.4%;	Pred. NO. 2.1e-05;		
Matches 535; Conservative	0;	Mismatches 1633;	Indels 22;	Gaps 6;

[illegible][illegible]

```

US-60-465-241-51800/c
: Sequence 51800. Application US/60465241
: GENERAL INFORMATION:
:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001468
: CURRENT APPLICATION NUMBER: US/60/465,241
: CURRENT FILING DATE: 2003-04-23
: NUMBER OF SEQ. ID NOS: 258418
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 51800
: LENGTH: 1202549
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(1202549)
: OTHER INFORMATION: n = A,T,C or G
: US-60-465-241-51800

```

[illegible]

Db	1093098	ATAAATATATATATGTTATATATATAAATATATATCTTATATATATATTTATATATATA	10930399
QY	1196	GGTCACATATACATATATTCCTTACAGAAAAACAACTTACAGAGAATTACATATCCA	1255
Db	1093038	AAATATATATATATATATATATATATATATAAATATATATATATATATATATATATATA	10929799
QY	1256	TATATGGGTATGCTATATACCTTTCACGATGCTATACTAGAGACTAAGAATAGTTATGTC	1315
Db	1092978	--AAATATATATATATATATATATATATATATAAATATATATATATATATATATAT	10929232
QY	1316	ATGCGATTAATAGAAATTCACAGCGGATATATATATGAGCCGATGTCGATCAC	1375
Db	1092921	ATTAAGATATATATATATATATATATATATATATATATAATATATATATATATATATA	1092862
QY	1376	TGCAAAATATATCTTGGTTGGTCACAAATATATAAACAACAAACAGAAAAAGAAAAACA	1435
Db	1092861	TATAAATATATATATATATATATATATATATATATATAAATATATATATATATATATA	1092802
QY	1436	TTTTTCTTGATTCATCAATGATGCTPAAAAATGCATAGATCTTTGGGCTTACAGTTTCA	1495
Db	1092801	TATAAATATATATAGTATATATATATAATATATATATATATATATATATATATAAATATACA	10927422
QY	1496	AGTCCCTACAGCGGTAAACATCTGCACTATATAATTCGTTCTTTAATGATCTTT	1555
Db	1092741	TTATATATATATAAATATATATATATATATATAAATATACATTATATATATAAATATATATAT	1092682
QY	1556	AACATATTTATTTTGTAGTGGAAATTTATATAAGACGAACTTGTACATTTACAAATATTTAT	1615
Db	1092681	AAAAATATACATTATATATATAAATATATATATATATATATATATATATAAATATATATTTTAT	1092622
QY	1616	ATTAGATACATAGTATGATTTATTCCAAAATACATACCTTGGATTTTAACTTAATCTG	1675
Db	1092621	ATAAATATATATAAATAT	1092562
QY	1676	TTTCTTCCTACGGTATATATATATATATCATGAGGTAAAAAAGTTTGTCTTATTTCCG	1735
Db	1092561	ATATAAATATACAAATATATAATATAAATAATATATATATATATATATATATATATAATATATA	10925020
QY	1736	GATGACGAAAGGATATAAACCCTPACATCTTATTTTGAAGAACTTAACCCCTTACAT	1795
Db	1092501	AAT	10924424
QY	1796	AGATTAATATACCGATGTGTTTGTGGCATATGACAGCCCTCTCAACCTGTGATAGTCAA	1855
Db	1092441	ATTATATACAAAT	10923822
QY	1856	TTTTTCTGCAATATTTAAATTAGCAATTCAATGCTCTCTATCAATACAGAAACAGCTGA	1915
Db	1092381	TAT	10923222
QY	1916	GATTAACATTTTATATAAGCAAAAAATTTTGAAGAAATGTTATATAATTTCTTACAAATAT	1975
Db	1092321	ATTCTAT	1092262
QY	1976	ATTAAATATATGATCCCTATATATGATATTTCCATGCTCTTATAAATATTTTATATATAT	2035
Db	1092261	ACTAT	10922022
QY	2036	AGTAT	2095
Db	1092201	TAT	10921422
QY	2096	TTTTGAATCTACAAATTTATATATATATAGCATATCAATGCATACAGAAAGTCCAAAA	2155
Db	1092141	ATATGACACAT	1092082
QY	2156	AAATTTGTTTACAGAAACTTCCAAATTTTTTTTTTTTTTATGAAACAAGAAATATACAGATA	2215
Db	1092081	ATTAT	1092022
QY	2216	GAAACATATTTTGTGGAATGGA-AGTATATATATATATATATATATATATATATATAT	2274

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 4854.18 Seconds
(without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202
Perfect score: 3202
Sequence: 1 atgtgggtattatattgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estcpl:*
7: em_estcro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195	6.1	473	AL768954	Arabidops
2	148	4.6	148	BH811044	SAUK_0571
3	146.4	4.6	148	AL768954	Arabidops
4	113.2	3.5	1101	CNS00EVL	AL069706 Drosophi
5	110.8	3.5	1101	CNS00EVL	AL069706 Drosophi
6	109.6	3.4	1101	CNS00213	AL061936 Drosophi

C	7	98.4	3.1	836	17	CNS01100	AL096642 Drosophi
C	8	98.2	3.1	1101	17	CNS000E07	AL069440 Drosophi
C	9	98	3.1	807	10	AV717372	AV717372 AV717372
C	10	98	3.1	1092	17	CNS020K7	AL175696 Tetradon
C	11	94	2.9	1101	17	CNS003BD	AL064091 Drosophi
C	12	94	2.9	1201	17	CNS0167M	AL063696 Drosophi
C	13	93.6	2.9	1101	17	CNS003BD	AL064091 Drosophi
C	14	93.6	2.9	1101	17	CNS000E07	AL069440 Drosophi
C	15	93	2.9	1101	17	CNS00B01	AL057419 Drosophi
C	16	92	2.9	661	17	CNS020VJ	AL209800 Tetradon
C	17	91.8	2.9	886	17	BH177277	BH177277 008_L_22-
C	18	91.8	2.9	886	17	CNS07JUX	AL161423 T3 end of
C	19	91.8	2.9	1101	17	CNS0039G	AL063092 Drosophi
C	20	90.6	2.8	945	17	CNS04D0K	AL285149 Tetradon
C	21	90.6	2.8	1187	17	B11102	B11102 F19C22-T7 I
C	22	90.4	2.8	836	17	CNS01100	AL099642 Drosophi
C	23	89.6	2.8	1029	17	CNS012GM	AL174271 Tetradon
C	24	89	2.8	987	17	CNS014PQ	AL104456 Drosophi
C	25	88.6	2.8	1092	17	CNS020K7	AL175696 Tetradon
C	26	88.6	2.8	1190	17	CNS020N7	AL206508 Tetradon
C	27	88.4	2.8	1225	17	CNS0161D	AL106171 Drosophi
C	28	88.2	2.8	1101	17	CNS0039G	AL063521 Drosophi
C	29	87.4	2.7	1169	17	CNS006KHQ	AL402900 T3 end of
C	30	87	2.7	1225	17	CNS0161D	AL106171 Drosophi
C	31	86.8	2.7	1101	17	CNS0021J	AL061936 Drosophi
C	32	86.6	2.7	866	17	AQ258984	AQ258984 nbx0021P
C	33	86	2.7	1013	17	CNS006RPQ	AL412260 T7 end of
C	34	85.8	2.7	1187	17	B11102	B11102 F19C22-T7 I
C	35	85.6	2.7	961	17	CNS00B01	AL051882 Drosophi
C	36	85.2	2.7	928	17	CNS00DXY	AL071865 Drosophi
C	37	85	2.7	1101	17	CNS05AB2	AL425825 Tetradon
C	38	84.8	2.6	1084	17	CNS071NH	AL425139 clone BA0
C	39	84.8	2.6	1101	17	CNS003BB	AL206508 Drosophi
C	40	84.8	2.6	1190	17	CNS020N7	AL206508 Drosophi
C	41	84.6	2.6	804	17	B12681	AL206508 Tetradon
C	42	84.4	2.6	1013	17	CNS006RPQ	AL412260 T7 end of
C	43	84.4	2.6	1101	17	CNS00B01	AL057419 Drosophi
C	44	83.8	2.6	609	17	CNS025K2	AL182171 Tetradon
C	45	83.8	2.6	886	17	BH177277	BH177277 008_L_22-

ALIGNMENTS

RESULT 1	AL768954	LOCUS	AL768954	473 bp	DNA	linear	GSS 18-JUN-2002
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,						
ACCESSION	AL768954	VERSION	AL768954.1	GI:21522073			
KEYWORDS	GSS.	SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana						
REFERENCE	1 Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H. and Weissshaar, B.						
AUTHORS	A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines						
TITLE	Unpublished						
JOURNAL	Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.						
REFERENCE	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)						
AUTHORS	for flanking sequence tag based reverse genetics						
TITLE	Unpublished						
JOURNAL	3 (bases 1 to 473)						
AUTHORS	Li, Y., Rosso, M., Strizhov, N. and Weissshaar, B.						
TITLE	Direct Submission						
JOURNAL	Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer						

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer Zuchtforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone T3A5. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: <http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
 source
 location/Qualifiers
 1. 473
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-082A08-011867"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 4.6%; Score 146.4; DB 17; Length 473;
Best Local Similarity 90.8%; Pred. No. 1.2e-11;
Matches 167; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1624 CTAGATGATGATATCCAAATACACTT-TGGATGTTTAACCTATCTGTTCTC 1682
 |||||
 DB 186 CTAGATGATGATATCCAAATACACTTGGATGTTTAACCTATCTGTTCTT 127
 QY 1683 CTACGCTATTAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGCGATCAT 1742
 |||||
 DB 126 CCTAGTATTAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGCGATCAT 67
 QY 1743 GAGGATTAACCTTAATTAATTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1802
 |||||
 DB 66 GAGGATTAACCTTAATTAATTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 7
 QY 1803 TTAC 1806
 |||||
 DB 6 TTAC 3

RESULT 4
CNS00EVL/C 1101 bp DNA linear GSS 04-JUN-1999
LOCUS CNS00EVL
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segefe@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseguwa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 location/Qualifiers
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR29B23"
 /clone_lib="RPCI-98"
 /note="end : 17"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 3.5%; Score 113.2; DB 17; Length 1101;
Best Local Similarity 35.8%; Pred. No. 4e-07;
Matches 232; Conservative 118; Mismatches 297; Indels 1; Gaps 1;

QY 1625 TAGTATGATGATATCCAAATACACTTGGATGTTTAACCTATCTGTTCTTCT 1684
 |||||
 DB 1100 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1041
 QY 1685 ACGGATTAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGCGATCAGA 1744
 |||||
 DB 1040 TATATATATTTTAAATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 981
 QY 1745 AGGATTAACCTTAATTAATTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1804
 |||||
 DB 980 TATATATTTTAAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 921
 QY 1805 ACCGATGTTTGTGCAATTAATGAGCGCTCTCACTGATGATGATGATGATGATG 1864
 |||||
 DB 920 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 861
 QY 1865 CAATATTAATTAATGAATTAATGCTACTATCAATAGAGAAACGCTAGATTAAT 1924
 |||||
 DB 860 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 801
 QY 1925 TTTAATTTAAGACAAATTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1984
 |||||
 DB 800 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 741
 QY 1985 TGATGCTATTAATGATTTTCCAT-TGCTTAATTAATTTTATTTTATTTTATTTA 2043
 |||||
 DB 740 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 681
 QY 2044 ATACATTTGAACCAATATAGTGTGTAATCAATATTCATTAATTTTGGAA 2103
 |||||
 DB 680 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 621
 QY 2104 TCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2163
 |||||
 DB 620 AAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 561
 QY 2164 TTAACAGAACCTCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2223
 |||||
 DB 560 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 501
 QY 2224 TTTTGTGGAATGAGTAGTATTAATTAATTAATTAATTAATTTTAA 2271
 |||||
 DB 500 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 453

RESULT 5
CNS00EVL

[illegible][illegible]

Db 1037 TTTT... 978
 Oy 1865 CAAATATTAATAGGAATTCATGCTATCATATAGAGAAACGCGATATACAT 1924
 Db 977 AAMATTAAMATTAATTTTCTTCAATTTTACATCTTACATCTTACATCTT 918
 Oy 1925 TTTAATTTAAGACAAATTTTGAAGAAATTTTGAAGAAATTTTGAAGAA 1984
 Db 917 TTTAATTTAAGACAAATTTTGAAGAAATTTTGAAGAAATTTTGAAGAA 858
 Oy 1985 TGAATCTTAAATTTTGAAGAAATTTTGAAGAAATTTTGAAGAAATTT 2044
 Db 857 AAAAAATTTT... 798
 Oy 2045 TACATTTAGAACCAATTAATAGTGTGGAATTTTGAAGAAATTTTGAAGAA 2104
 Db 797 AAAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 741
 Oy 2105 CTACAAATTTAATTAATTTAGTCAATTAATTAATTAATTAATTAATTA 2164
 Db 740 TTTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 681
 Oy 2165 TTAACGAAATCTCAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2224
 Db 680 TTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 621
 Oy 2225 TTTGTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAAT 2281
 Db 620 TTTT... 564
 RESULT 7
 CNS01100/c
 LOCUS 836 bp DNA linear GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
 fly), genomic survey sequence.
 ACCESSION AL069642
 VERSION AL069642.1 GI:5611253
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 836)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 COMMENT - Web : www.genoscope.cns.fr)
 determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (BDGP) -
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.
 FEATURES
 Location/Qualifiers
 source 1..836
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN05N18"
 /clone_1ib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : 77"
 BASE COUNT 379 a 58 c 42 g 269 t 88 others
 ORIGIN
 Query Match 3.1%; Score 98.4; DB 17; Length 836;
 Best Local Similarity 45.3%; Pred. No. 5.8e-05;

Matches 264; Conservative 24; Mismatches 290; Indels 5; Gaps 1;
 Oy 1546 ATGCATCTTAAATATTTATTTAGTGGAAATTTAAGAGGAACTTACATTA 1605
 Db 592 ATCTTNNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNN 533
 Oy 1606 CAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1665
 Db 532 TTAATAATAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 473
 Oy 1666 CTAAATCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1725
 Db 472 TGAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 413
 Oy 1726 T-----TATTTGGGATTCAGAGGATTAACCTTAATTTATTTATTT 1780
 Db 412 TTTAATTTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 353
 Oy 1781 AACCTTTTACATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 1840
 Db 352 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 293
 Oy 1841 AACTGTGATGCAATTTTCTGCAATTTTATTTATTTATTTATTTATTT 1900
 Db 292 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 233
 Oy 1901 AGAAGAAACGCTGATTTATTTATTTATTTATTTATTTATTTATTTAT 1960
 Db 232 AATTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 173
 Oy 1961 ATTTCTAACAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2020
 Db 172 TTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 113
 Oy 2021 TTTT... 2080
 Db 112 AATTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 53
 Oy 2081 ATCTCATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2123
 Db 52 ATTTT... 10
 RESULT 8
 CNS00E07/c
 LOCUS 1101 bp DNA linear GSS 04-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR29P01 of RPCR-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069440
 VERSION AL069440.1 GI:4949583
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 COMMENT - Web : www.genoscope.cns.fr)
 determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see: http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Oosogawa and
 Aaron Mammosier in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCR-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE

1. 1101
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_1lb="RPCI-98"
/note="end: TERT3"
BASE COUNT 366 a 66 c 104 g 351 t 214 others
ORIGIN

Query Match 3.1%; Score 98.2; DB 17; Length 1101;
Best Local Similarity 41.0%; Pred. No. 5.3e-05;
Matches 229; Conservative 77; Mismatches 237; Indels 15; Gaps 2;

QY 1786 TTTTACTCATGATTAATTCACGTAATGTTTGTGCGCATATGACAGCCTCTACAACTG 1845
D 1031 TTTTMMAMAAACAMAAATTTTMTATTTTTTTTTTMAATATTCMATATTTMAATACA 972
QY 1846 TGATAGCAATTTTTCGCAATATTAATAGGAATTCATGCTACTACATACAGAG 1905
D 971 WMAVATTTTWTATCATATATTTTATATACAAATTTMAAATAAATAACWMAAT 912
QY 1906 AAACAGCTGAGTA-TTACATTTTAAATTAAGACAAATTTTGAATAATTTTAAATTT 1964
D 911 TTAATAAACATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 852
QY 1965 CTACATATTTATTAATATGATGCTATATATTTTCTATGTTCTTAAATATTTT 2024
D 851 AAAAATAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 792
QY 2025 TTTTATATTTAGTATTAATATACATTAATTAATTAATTAATTAATTAATTAATTA 2084
D 791 WMAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 732
QY 2085 CCAATTAATTTTGAATCTACAAATTAATTAATTAATTAATTAATTAATTAATTA 2144
D 731 TTTTATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTA 672
QY 2145 AGTTCACAAAATAATTTGTTAACAGA-----AATTCGCAATTTTCTTTT 2190
D 671 WTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 612
QY 2191 TTTATGACAAAGAAATACAGATAGAAACTATTTGTTGGAATGGAAGTAGATATA 2250
D 611 WMCATTAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 552
QY 2251 TACATTAAGCAATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2310
D 551 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 492
QY 2311 GTAGGTGAATTAATTAAT 2328
D 491 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 474

RESULT 9
AV717372/c 807 bp mRNA linear EST 16-Oct-2000

LOCUS AV717372 DCB Homo sapiens cDNA clone DCBCH07 5', mRNA sequence.

DEFINITION AV717372
ACCESSION AV717372
VERSION AV717372.1 GI:10814524

KEYWORDS
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 807)
Xu, X., Gu, J., Liu, F., Qu, J., Zhao, M., Li, Y., Huang, Q., Zhou, J.,

TITLE
JOURNAL
COMMENT
Contact: Zenguan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzge@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

SOURCE

1. 807
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DCBCH07"
/clone_1lb="DCB"
/cell_type="dendritic cells"
/dev_stage="mature"
/lab_host="BM25.8"
/note="Vector: pTrioEx2; Site_1: sfIIA; Site_2: sfIIB"
BASE COUNT 388 a 50 c 20 g 345 t 4 others
ORIGIN

Query Match 3.1%; Score 98; DB 10; Length 807;
Best Local Similarity 46.2%; Pred. No. 6.7e-05;
Matches 323; Conservative 0; Mismatches 376; Indels 0; Gaps 0;

QY 1577 AATTATTAAGACGCACTTTGATTTTAACTTAATCTGTTCTCCACGTAATTAATTA 1636
D 773 AATTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 714
QY 1637 ATCCAAATACATCTTTGATTTTAACTTAATCTGTTCTCCACGTAATTAATTA 1696
D 713 AATTATTAATTTTAAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 654
QY 1697 TTAATCATCGAGTAAGAAAGTTTGTCTTATTTTTCGCGATGATGAAGATAAACCCTA 1756
D 653 AAAAATATATTAATTAATTAATTTTGTCTTATTAATTAATTAATTAATTAATTAATTA 594
QY 1757 ATGACTTAATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1816
D 593 TTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 534
QY 1817 TGTTCATTAATGACAGCTCTACACGTAATTAATTAATTTTCTCAATTAATTAATTA 1876
D 533 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 474
QY 1877 TAGGAATTCAGTCTACTATCAATAGAAAGAAAGCTGAGTATTAATTAATTAATTAAG 1936
D 473 TTGATTTTAAAGTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 414
QY 1937 ACAAAATTTTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1996
D 413 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 354
QY 1997 TGTATTTCTATGTTCTTAAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTA 2056
D 353 ATTAAGCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 294
QY 2057 CAATTAATGTTGGAATTCAAATATCTCATTAATTTTGAATTAATTAATTAATTAATTA 2116
D 293 CTTTCTTAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 234
QY 2117 AATATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2176
D 233 AATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 174
QY 2177 CCAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2236
D 173 GTTACTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 114

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 2933.8 seconds
(without alignments)
10921.753 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcaacaattattaat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vi:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vi:*

30: em.htg.hum:*

31: em.htg.inv:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pln:*

35: em.htg.fod:*

36: em.htg.mam:*

37: em.htg.vrt:*

38: em.sy:*

39: em.htgo.hum:*

40: em.htgo.mus:*

41: em.htgo.other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	4818	8 AF044216	AF044216 Arabidops
2	1090	99.0	84196	8 ATW3A5	AL132979 Arabidops
3	62.6	5.7	4629	3 AF273674	AF273674 Plasmodi
4	61.4	5.6	12029	3 AE001372	AE001372 Plasmodi
5	60.8	5.5	129360	2 AC117079	AC117079 Dictyoste
6	60.6	5.5	3364	14 TSP418778	AJ418778 Tomato sp
7	60.6	5.5	171317	9 AC020941	AC020941 Homo sapi
8	60	5.4	233269	2 AC094291	AC094291 Rattus no
9	59.4	5.4	224635	2 AL732314	AL732314 Homo sapi
10	59	5.4	253305	3 PFMAL3P7	AL034559 Plasmodi
11	58.6	5.3	163669	2 AC114238	AC114238 Rattus no
12	58.6	5.3	192389	9 AC007182	AC007182 Homo sapi
13	58	5.3	115489	2 AC117072	AC117072 Dictyoste
14	57.8	5.2	27291	2 AC115575	AC115575 Dictyoste
15	57.8	5.2	169546	2 AC004157	AC004157 Plasmodi
16	57.6	5.2	35077	3 CEY388A	AL021483 Caenorhab
17	57.6	5.2	144784	9 AC009224	AC009224 Homo sapi
18	57.6	5.2	188993	2 AC109656	AC109656 Rattus no
19	57.2	5.2	18701	2 AC116985	AC116985 Dictyoste
20	57.2	5.2	73666	9 HSJ333B15	AL109954 Human DNA
21	57	5.2	43666	3 CEY51H1A	AL032644 Caenorhab
22	57	5.2	104992	2 AC005504	AC005504 Plasmodi
23	57	5.2	156060	2 AC004153	AC004153 Plasmodi
24	56.8	5.2	48811	2 AC116550	AC116550 Dictyoste
25	56.8	5.2	160759	2 AC117082	AC117082 Dictyoste
26	56.6	5.1	6107	6 AX252043	AX252043 Sequence
27	56.6	5.1	6107	6 AX344427	AX344427 Sequence
28	56.6	5.1	6107	6 AX348822	AX348822 Sequence
29	56.6	5.1	56099	2 AC115598	AC115598 Dictyoste
30	56.6	5.1	67970	3 PFMAL1P3	AL031746 Plasmodi
31	56.6	5.1	86515	2 AC127730	AC127730 Rattus no
32	56.6	5.1	159941	2 AC016220	AC016220 Homo sapi
33	56.6	5.1	178642	9 AC023122	AC023122 Homo sapi
34	56.4	5.1	65691	3 PFMAL3P1	297348 Plasmodi
35	56.4	5.1	99003	2 AL390756	AL390756 Homo sapi
36	56.4	5.1	110000	2 PFMAL13P2_2	Continuation (3 of
37	56.4	5.1	124041	9 CNS01DSJ	AL121808 Human chr
38	56.4	5.1	183353	9 AC087071	AC087071 Homo sapi
39	56.4	5.1	185273	9 AC073320	AC073320 Homo sapi
40	56.2	5.1	45296	2 AC115613	AC115613 Dictyoste
41	56.2	5.1	163443	2 AC006280	AC006280 Plasmodi
42	56.2	5.1	196149	2 AC004709	AC004709 Plasmodi
43	56	5.1	204439	9 AP000807	AP000807 Homo sapi
44	55.8	5.1	2166	3 AF200327	AF200327 Plasmodi
45	55.8	5.1	14867	3 AE001398	AE001398 Plasmodi

ALIGNMENTS

RESULT 1

AF044216

LOCUS

DEFINITION

Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,

complete cds.

ACCESSION

AF044216

VERSION

AF044216.1 GI:2935341

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana.

Eukaryota; Viridiplantae;

Spermatophyta; Magnoliophyta;

Rosidae; eurosids II; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

4818 bp DNA linear PLN 25-JUN-2001
Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,

complete cds.

AF044216

AF044216.1 GI:2935341

Arabidopsis thaliana.

Eukaryota; Viridiplantae;

Spermatophyta; Magnoliophyta;

Rosidae; eurosids II; Brassicaceae; Arabidopsis.

1 (bases 1 to 4818)

Choe,S., Dilkes,B.P., Fujioka,S., Takatsuto,S., Sakurai,A. and

Pred. No. is the number of results predicted by chance to have a

Feldmann, K.A.
The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 22alpha-hydroxylation steps in brassinosteroid biosynthesis
Plant Cell 10 (2), 231-243 (1998)
98158690
9490746
2 (bases 1 to 4818)
Choe, S., Dilkes, B.P., Aspiroz, R. and Feldmann, K.A.
Direct Submission
Submitted (22-JAN-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
Location/Qualifiers
1. 4818
/organism="Arabidopsis thaliana"
/strain="Ws-2"
/db_xref="taxon:3702"
/chromosome="III"
/map="8 cm from nga6"
990. 4398
/gene="DWF4"
/note="dwarf4"
/allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"
990. 1055
/gene="DWF4"
Join(<1133, 1353, 1434, 1758, 1844, 1996, 2095, 2409, 2562, 2654, 2746, 2824, 2931, 3040, 3795, 4398)
/gene="DWF4"
/product="steroid 22-alpha-hydroxylase"
Join(1133, 1353, 1434, 1758, 1844, 1996, 2095, 2409, 2562, 2654, 2746, 2824, 2931, 3040, 3795, 4040)
/gene="DWF4"
/function="brassinosteroid biosynthetic enzyme"
/note="member of the cytochrome P450 superfamily; CYP90B1"
/codon_start=1
/product="steroid 22-alpha-hydroxylase"
/protein_id="AAC05093.1"
/db_xref="GI:2935342"
/translation="MFETEHTLLPLLLPSLLSLLFLILKRRNRKTRFNLPPGKS
GWPFLEGTIGLYNATATLGTGDMQHVSKYKIVRSNLFGEPTIVSDAGLNRFILO
NEGRLEFCSYPRSIGILGKMSLVLDGMDHRDMSISINLSHARLRTILLKDYERH
TLFVLDNQNSIFSAQDEAKKFTENLMAKHMSMDPGEETEQLKKEVYFPMKGVVS
APLNQPTGATYHALQSRATILKFIERKMEERKLDKEEDDEEEVKTSDEAEMSKSDH
VRKQTDLLGLYHLKSNLSTFEQILDILSLFAGHETSSVAIALAIFLQACPKAV
EELREHLEIARAKELGESELDNDYDKMDFTQCVINETLRLNVVFLHRKALKDV
RYKGYDIPSGNWKVLPVISAHLNDSRYDQNLNFWRWQONNGASSSGSFSFTWGN
NMPYGGGPRLCAGSELAKLEMAVFIHLVLKFNWELAEDDQPFAPFVDFPNGLPIR
VSRIL"

BASE COUNT 1566 a 727 c 888 g 1637 t
ORIGIN

Query Match 100.0%; Score 1101; DB 8; Length 4818;
Best Local Similarity 100.0%; Pred. No. 4.3e-180;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTACAAATATTATTTAGTCAATAAATGATAGAGTTCACAAATTTT 60
DB 32 AATCTACAAATATTATTTAGTCAATAAATGATAGAGTTCACAAATTTT 91
QY 61 TGTAAACAGAACTTCCAAATTTTATTTATGGAACAGAAATACAGATAGAAAC 120
DB 92 TGTAAACAGAACTTCCAAATTTTATTTATGGAACAGAAATACAGATAGAAAC 151
QY 121 TATTTTGTGTGGAATGGAATGATATATACATTAAGCAAAATTTTAAAAATTTATATAA 180
DB 152 TATTTTGTGTGGAATGGAATGATATATACATTAAGCAAAATTTTAAAAATTTATATAA 211
QY 181 GCCTATACGGCTCAAAGTATGTTATCTAGTAGGTGTTAATTAATGATGCGGATTT 240
DB 212 GCCTATACGGCTCAAAGTATGTTATCTAGTAGGTGTTAATTAATGATGCGGATTT 271
QY 241 CAGAAATTCGACAAACATGAAACCGAATTAATAATTAACCTTTAAATAATAATAAATT 300
|||||

Db 272 CAGAAATTCGACAAACAAATGAAACGGAATTAATAATTAATAATAATAAATT 331
QY 301 TGAGTAAATGTGTTTTCTGACTATTTCAGGGGCAAAAAAGACAAATGCCAAAGTCTACG 360
Db 332 TGAGTAAATGTGTTTTCTGACTATTTCAGGGGCAAAAAAGACAAATGCCAAAGTCTACG 391
QY 361 GGTTCGACTGTCCAGTTCGGTAATAATCAATAAATCTGCTTTGACCGCAGCTCGTGT 420
Db 392 GGTTCGACTGTCCAGTTCGGTAATAATCAATAAATCTGCTTTGACCGCAGCTCGTGT 451
QY 421 AGGGTCTCTGACATTTTCTGCTTACCCCTACTCGTGAGCCACCCCTTTCCCAT 480
Db 452 AGGGTCTCTGACATTTTCTGCTTACCCCTACTCGTGAGCCACCCCTTTCCCAT 511
QY 481 ATCCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGACTTCCT 540
Db 512 ATCCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGACTTCCT 571
QY 541 GGGATTCTCTGGAGCATTTTATCAAAAAATTTATAGCAGCAATGGGTTTATTAATTTAAA 600
Db 572 GGGATTCTCTGGAGCATTTTATCAAAAAATTTATAGCAGCAATGGGTTTATTAATTTAAA 631
QY 601 ACTCACAACCTGATCAGATAAAATTTCAATAAACACTTTTACGATGATTCGACGATCA 660
Db 632 ACTCACAACCTGATCAGATAAAATTTCAATAAACACTTTTACGATGATTCGACGATCA 691
QY 661 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 720
Db 692 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 751
QY 721 CAATTGATTATAGATATATCCATTATTCATCATGATATTTATGATATAAATAGCTTTAAAC 780
Db 752 CAATTGATTATAGATATATCCATTATTCATCATGATATTTATGATATAAATAGCTTTAAAC 811
QY 781 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTTTTAAATTAAGAGTTTAAATAATAA 840
Db 812 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTTTTAAATTAAGAGTTTAAATAATAA 871
QY 841 AGTATTAAGAGGATTAACGAGGCAACAAAAGTAAATGAACGAGAGAAACAAAGCCAT 900
Db 872 AGTATTAAGAGGATTAACGAGGCAACAAAAGTAAATGAACGAGAGAAACAAAGCCAT 931
QY 901 GAAGTCAATGTTTGTAGTTTAAAGCAATTAAGAGAGTTTATTAATAATTTAAATGACGATGA 960
Db 932 GAAGTCAATGTTTGTAGTTTAAAGCAATTAAGAGAGTTTATTAATAATTTAAATGACGATGA 991
QY 961 TAACAATATATTTCTGACTTTTAAACCCCTCTTACAAACAGAGAGCTCCCTTTT 1020
Db 992 TAACAATATATTTCTGACTTTTAAACCCCTCTTACAAACAGAGAGCTCCCTTTT 1051
QY 1021 CAGTAGAGTCCGATTCCTCAATCTTAAAGACAAAGCCATTAGAAAGAGAGAGTGAAGTGA 1080
Db 1052 CAGTAGAGTCCGATTCCTCAATCTTAAAGACAAAGCCATTAGAAAGAGAGAGTGAAGTGA 1111
QY 1081 AGAGAGAGAGAACTAGCTCC 1101
Db 1112 AGAGAGAGAGAACTAGCTCC 1132

RESULT 2
ATT3A5/c
LOCUS
DEFINITION
Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
ACCESSION
AL132979
VERSION
AL132979.2 GI:6782244
KEYWORDS
Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
1 (bases 1 to 84196)
REFERENCE
Bloeker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F. and


```

/number=6
complement(28374. .28698)
/number=7
complement(28699. .28778)
/number=7
complement(28779. .28999)
/number=8
38307. .41323
/genes="T3A5.50"
complement(join(38307. .38796,38877. .39090,39189. .39270,
40180. .40242,40337. .40401,40489. .40564,40655. .40734,
40805. .40919,41225. .41323))
/notes="Contains eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
/codon_start=1
/product="U1 snRNP 70K protein"
/protein_id="CAB62436.1"
/db_xref="GI:6561970"
/translation="MGDSGDFLRNPNAVQAKVQNRANVQLKMGQSHPTGLTN
NLKLFEPRLPLEYKPPPEKRCPPYGMQAFVSFNAEPGDPEYAPPPPELPSQKR
ERIKRLKEGVKAEADLKRYDPNDPNATGDKYTLFVSLNLTSSSESKIKREFES
YGPVKRLVLDLTNPKGVAFIEYMTKMDKAAV KQADGOKIDGRVLDVVERGT
VNPRLRGLGLGTSRVGGEEIVGEOQPOGRTSQSEPSRPREKRSREKGRERE
RSRELSHQOPERSDRPREDKHHRDQGGDRDRDRDRDRDRDRDRDRDRDRDR
GDRTSRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR
TDRYSHRDKMEEDDFRYREYKRSKRSREYVR"
complement(38307. .38796)
/number=1
complement(38797. .38876)
/number=1
complement(38877. .39090)
/number=2
complement(39091. .39188)
/number=2
complement(39189. .39270)
/number=3
complement(39271. .40179)
/number=3
complement(40180. .40242)
/number=4
complement(40243. .40336)
/number=4

```

Query Match 99.0%; Score 1090; DB 8; Length 84196;
 Best Local Similarity 99.9%; Pred. No. 1.9e-178;
 Matches 1101; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 AATCTACAAATATTAAATATTAGTCAATAACAATGCATAGAAAGTTCCAAAAAATTT 60
DB 30101 AATCTACAAATATTAAATATTAGTCAATAACAATGCATAGAAAGTTCCAAAAAATTT 30042
QY 61 TGTTAACAGAACTTCCAAATTTTATGGAACAAGAAATAACAGATAGAAAC 120
DB 30041 TGTTAACAGAACTTCCAAATTTTATGGAACAAGAAATAACAGATAGAAAC 29982
QY 131 TATTTTGTGTGGAATGGAAGTAGTAATACATTAAGCAAAATTTAAAAAATATATAA 180
DB 29981 TATTTTGTGTGGAATGGAAGTAGTAATACATTAAGCAAAATTTAAAAAATATATAA 29922
QY 181 GCCTATACGCGCTCAAGTATGTTATCTAGTAGGTGTAATTAATATCATGCTGGGATT 240
DB 29921 GCCTATACGCGCTCAAGTATGTTATCTAGTAGGTGTAATTAATATCATGCTGGGATT 29862
QY 241 CAGAAATGGGCAACAATGAAACGGAATTTAAATATTAACTTTAAATATAAATAAATTT 300
DB 29861 CAGAAATGGGCAACAATGAAACGGAATTTAAATATTAACTTTAAATATAAATAAATTT 29802
QY 301 TGAGTAAATGTGTTCTTCTGACTATTGAGGGGCAAAAAAAGACAATGCCAAAGTCTACG 360
DB 29801 TGAGTAAATGTGTTCTTCTGACTATTGAGGGGCAAAAAAAGACAATGCCAAAGTCTACG 29742
QY 361 GGTGTGACTGTCCAGTTCGGTAATTAATCTAATCTGCTTTACCGCGCGCTCGTGT 420
DB 29742 GGTGTGACTGTCCAGTTCGGTAATTAATCTAATCTGCTTTACCGCGCGCTCGTGT 29682

```

```

Db 29741 GGTGTGACTGTCCAGTTCGGTAATTAATCTAATCTGCTTTGACCGCACGCTCGTGT 29682
QY 421 AGGGGTCTCTTGACATTTTCACTGTCTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 480
DB 29681 AGGGGTCTCTTGACATTTTCACTGTCTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 29622
QY 481 ATCTTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCCT 540
DB 29621 ATCTTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCCT 29562
QY 541 GGGATTCTCTGGAGCATTATCAAAAATTTATAGCACGAATGGGTTTATTAATTTAAAA 600
DB 29561 GGGATTCTCTGGAGCATTATCAAAAATTTATAGCACGAATGGGTTTATTAATTTAAAA 29502
QY 601 ACTCACAACTTGCATCAGATATAAATTTCAATAAACACATTTTACGATGATTTCGTACGATCTA 660
DB 29501 ACTCACAACTTGCATCAGATATAAATTTCAATAAACACATTTTACGATGATTTCGTACGATCTA 29442
QY 661 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 720
DB 29441 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 29382
QY 721 CAATTGATTATAGATATATCCATTATCCATGATATTATGATATAAATAGCTGTTAAAC 780
DB 29381 CAATTGATTATAGATATATCCATTATCCATGATATTATGATATAAATAGCTGTTAAAC 29322
QY 781 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTTGTTTAAATTTAAAGAGTTTAAATAATAA 840
DB 29321 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTTGTTTAAATTTAAAGAGTTTAAATAATAA 29262
QY 841 AGTATTAAAGGAGCATAACGAGGCAACAAAAGTAAATGAACGAGGAGAAACAAAGCCAT 900
DB 29261 AGTATTAAAGGAGCATAACGAGGCAACAAAAGTAAATGAACGAGGAGAAACAAAGCCAT 29202
QY 901 GAAGCTCATGCTTA-GTTTAAAGCTTAATAAGAGATTTTATTAATTTTAAATGACGATG 959
DB 29201 GAAGCTCATGCTTAATAAGAGATTTTAAAGAGATTTTATTAATTTTAAATGACGATG 29142
QY 960 ATAACAATTTATTTTCTGACTTTCTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 1019
DB 29141 ATAACAATTTATTTTCTGACTTTCTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 29082
QY 1020 TCAGTAGAAGTCCGATTTCCCAATCTTAAAGACAAACCAATTTAGAAAGAGAAAGTGTGATGA 1079
DB 29081 TCAGTAGAAGTCCGATTTCCCAATCTTAAAGACAAACCAATTTAGAAAGAGAAAGTGTGATGA 29022
QY 1080 GAGAGAGAGAGAACTAGCTCC 1101
DB 29021 GAGAGAGAGAGAACTAGCTCC 29000

```

RESULT 3

AF273674

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (31-MAY-2000)

Seattle Biomedical Research Institute, 4

AF273674 4629 bp DNA linear INV 24-APR-2001
 Plasmodium falciparum DNA-dependent RNA polymerase (TRNA) gene,
 complete cds; nuclear gene for probable mitochondrial product.

AF273674
 AF273674.1 GI:9857984

Plasmodium falciparum.
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 4629)

Li J., Muga, J. A., Cermakian, N., Cedergren, R. and Feagin, J. E.
 Identification and characterization of a Plasmodium falciparum RNA
 polymerase gene with similarity to mitochondrial RNA polymerases
 Mol. Biochem. Parasitol. 113 (2), 261-269 (2001)

21192559
 11295180
 2 (bases 1 to 4629)
 Li J., Muga, J. A., Cermakian, N., Cedergren, R. and Feagin, J. E.

Direct Submission
 Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4

FEATURES	source
Nickerson St., Seattle, WA 98109-1651, USA	
Location/Qualifiers	
1. .4629	
/organism="Plasmodium falciparum"	
/db_xref="taxon:5833"	
/chromosome="11"	
/clone="C10"	
/note="West African origin"	
<70. .>4581	
/gene="TRNAP"	
<70. .>4581	
/gene="TRNAP"	
/product="DNA-dependent RNA polymerase"	
70. .4581	
/gene="TRNAP"	
/codon_start=1	
/product="DNA-dependent RNA polymerase"	
/protein_id="AAG00950.1"	
/db_xref="GI:9857985"	
/translation="MFVCKKCSRYVKKIPIFYICESKYQLFLKRGNDYKNIFLEK	
GYISTHVGDKKTKQKKNDLDRFLDKPSTNIDKDEYKTKDNNNEEYNICN	
NRLINNNKINSMMHTNFRNKKVCVVDNNINNTINNTYFEQKEKEEYINTPKAK	
NLNLPYKKNKDIIDKKLDQVVKLHNLISIKLIDKDKENIKIYINTLUNKY	
VKNKNIYFFENLDIYMSVGLLIDIGGEIFNYINKLNNIDDKETNDYD	
EKVHNTNFRFLMDVVDKDNINNNEDINYIEKKRIYIGNIPIYVSKSEKE	
VPPKKKNNNNIIVNNNINNNKSGDYNINCTNDLGEKTKHNTYNNPKFDY	
NEEKEEKEESINVRQMLIRSSYKKAIEAEFEVSNLHDKTKITEGGLCKY	
LWLTSELEKILYIKKNNKKKIDPFDIDKLLATITVKTQYTNPLKKNYSN	
NVSTQSYFHEIYSYSLTHAIKIGFEINNELNFKLKKHLYNNKKNNSSF	
SHQYKRIKQKQQTGKGNHQMHLGGDNOTSHLNCANNEDINSANNFNIDN	
SCANNFGNSCNGNFGNSDVEYGHNTNHHVHKDDDDIENWSVKKATIGGLL	
NMLTENAKMDVIDNAKNEFKYKYLFLKTKANNAEDLYECKYKKELYE	
EESTNLKFIILDKYKNNSDYNNFKYMEKIKIYNTGTHDKNNVKNQNNQNGD	
KHNCANROELQSDNKKESFCOKREKLMEKILYRSKDRNKKILPFIHRYVK	
NNWYGVHRECCANYLNNAINSHIPLNLPMPCKRWENFEGMLLKNSELR	
NIKPLNLYNCLNNKNIIVNSIGVWKINKETILHYIEVYIHTGIIKPIPLYKNT	
LPKINLKQNNEEKYLLKLEINRLNKLISERTFLQKLAVAKTKPDNDIYFP	
HNDFPQNNIINIKLTQDFNIEFWNAEKPMQALAYADLNKNALESFNASKYK	
FSPIQDQCTGLQYAAAGKYGKGVANNIPSDPEQDIYSVVDIVISKINDLM	
NISNGHNNIISFENIKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	
KFKDLKRVKQKQITMICYGTSIGAKNOVKQIQTMIGKELDKNINIKLSOYIAN	
IPESISEPKRAMIKKFWNLNLSKVTNELNIPITWISPIGLCEQYRLGNILYNT	
LOSISVYSKQNLHKNRGLGFPFNFVSHLDASHLMATAEKWVIONNFSFAVHDSY	
WHAQNVDMQOITRESFITYLNPENILEYNNFRILGKFAKIPSPSQGELDISM	
VKDSLYFFS"	
BASE COUNT: 2114 a 481 c 621 g 1413 t	
ORIGIN	
Query Match	5.7%; Score 62.6; DB 3; Length 4629;
Best Local Similarity	48.2%; Pred. No. 0.16;
Matches	176; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY	11 TTATTAATATTTAGTCAATAACATGATAGAAAGTTCCAAAAAATTTCTTAAACAGA 70
DB	
2123 TTATTGAGATGCAAAATGGATGCTGATATTATGTCAGGAAATGAATTTAAAAATG 2182	
QY	71 AACTTCCAAATTTTTTTTTTTATGGAACAAGAAATTAACAGATAGAAAACTATTTGTTG 130
DB	
2183 ATATAATAATATTTATTTTATTTATACATAAAACAGCCAAAAATAATCAGAGGATTTATG 2242	
QY	131 TGGAAATGGAAGTAGTATATACATTAAGCAAAATTTTAAAAATTTATATAAGCCTATACGC 190
DB	
2243 AAGATGTGTAATAATAAAAGAAATTAATATCAAAAATTTCTATGAGAGAAATCCCAATA 2302	
QY	191 GCTCAAGTATGTTACTAGTGTAAATTAATATGATGCGTGGCATTTCCAGAAATTTGGG 250
DB	
2303 CATTAATAATTTATTTCTTTGATTAATAATAAAAGAAATAATAGTATGATATAATTTTA 2362	
QY	251 ACAACAATGAAACGGAATTAATAATTAACCTTTAAATAATAATAAATTTAGTAAATG 310
DB	
2363 AAAAAAATTTATGGAATAAAATAATGATACCTAATAATTTGGAACACATGATAAAATG 2422	


```

WYCSCHRYNLSPHHLXVCMKQKQVGLKKNLPPIYSNTEFSSRSYFALLIFMGTLIS
MMLPSTLPQONSXIYGVDTILNNHNNNNNGVSGSGSGTGGATVNSLLNFNF
NGILSFSNEVSGGMFMNPLVINGNNTTACYNPGLNPSVSYISFYFSSSPS
IYSSGSSSSAQLNFHQHYVNSIQSNNNKNNNNNDNNNNINIKKNGKDLNE
KLFSDFDIYFNSDSKLNQDEKFSFVCHSLSYSPNYLNNNNPNQOQV
NPHFTPIVRSSTSGGSSNNDESDTSISQLKQKHVLLKNYFNTHINENYOPT
PTTSTTLQNGENCIDENLOQOQETEQOQSKYSTKSPILNLSGSEYHDDIH
LQNIYIAHQHVRVTIADSKNFICTLIPVRKFCY"
/translation="MNMNLIPVNSFSDNSGTYTYTIGNYVDVIEITSIARGPNTIDI
KFNPSIKNSKIVLLDGRNLOPQNFYNTVTIILHINISPLSGSUSLNKNTYKLD
GNFHPLESGITPRPYSHLIFIIYKWSFLVILKILKLNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
join(13528..13795,13916..14022)
/notes="ORF_ID:dd_02320"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43691.1"
/db_xref="GI:21166073"
/translation="MHMNLIPVNSFSDNSGTYTYTIGNYVDVIEITSIARGPNTIDI
KFNPSIKNSKIVLLDGRNLOPQNFYNTVTIILHINISPLSGSUSLNKNTYKLD
GNFHPLESGITPRPYSHLIFIIYKWSFLVILKILKLNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
join(13528..13795,13916..14022)
/notes="ORF_ID:dd_02320"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43691.1"
/db_xref="GI:21166074"
/translation="MNIDVFKLAKGFFGRSKNCKSLARVERKGLNYYVRKLKRR
DFRKMVIERINAGARHEMTYSDFQGLIASGVEINRVLVLSIGAVSEPSFKVLVD
HSKNTVSNLHKSYIPLDQPN"
/complement(join(14375..14423,14526..14851))
/notes="ORF_ID:dd_02323"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43692.1"
/db_xref="GI:21166075"
/translation="MESKOMEIPKEDPWGLWKAEMKWGWGVLDNDMLKQGLEERTK
LADHLLKKEKEKEAIGSGSL"
/complement(join(15910..15915,16125..16229,16310..16392,
16502..16576,16692..16710))
/notes="ORF_ID:dd_02326"
/codon_start=1
/product="vegetative specific protein H7"
/protein_id="AAM43693.1"
/db_xref="GI:21166076"
/translation="MDVQTYGAGQNVKLGANOKKIVESEEDIALPELNPSPQALQ
RARNALMTORELAFINERPGVINEYESAIPSOAVLSKLEKALNVKLN"
join(17001..17789,17899..18159)
/notes="ORF_ID:dd_02308"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43694.1"
/db_xref="GI:21166077"
/translation="MSSYIDNINNNNNNNKIINIYILNNDCLVMDVGDILIRTKYRL
LGSIVGSEVNLRLNSTINSVPLIISPYEIRLGLEKOWFOIINESATPPTIDHLSLF
KESRQTYQSLMDKYKQRLKKEKPRKLNNTNSNKNKNIPLSDTDSNNNNN
TNBENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
EKEEINESKIFPNVDISIYCSKEALDNEPIKTLWRDGRDYITILSRGYLSGGS
KFGGTFLAYGKEFPLPATFIVVIKRSNQTFQSLDIITARLAVHVNKTTLIASLDEE
TNPIYIANKWGS"
join(19303..20698,21044..21077,21937..22371,22484..24078,
24154..26522)
/notes="ORF_ID:dd_03155"
/codon_start=1
/product="Opaque-specific ABC transporter CDR3."
/protein_id="AAM43695.1"
/db_xref="GI:21166078"
/translation="MSTIKISSEFFKSTFSTINGKDKSINLNNDNKNNNNNNNNN
NNNNNNNNNDNDIKPTGGIDMTLITATSLTIPTKGGDIIESDTLKSMLIIDDT
NDNSSTTTVGGDNNNEQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
WPTWKDYENRYDLRIFISRNGLDTISRLNSGFGNFVYQNTFKFSKLYWKNSIF
GDLEVPDNNNDIITASPTSGIKQLNITSYNGVINEEDDDRLGLIYVWKNSIF
PGCVHGLASLMDSEMACCIRYFNNFDDNLQKQKQKQOQOQOQOQOQOQOQOQOQ
VIPPITNSKTYVTANLSVNYKFFVLESFVSVETRIQRDKGKIFISITIKDNGVI
RAESTAIIWVKIDTPISTISSPSPSLPSPSTSNKLNENNNNNNNNNNNNNNN

```

```

NNNNNNIQQDNNNSCKHSYDTIDMLSISENNNNNNNNNNNNNNNNNNNNNNNNNDY
DYDSINNEEFENYSKEGSIKFRIDGKNNNNHDIELGKPNEDFKLROY
FEDSORKMLNHNKPKMGVSTKNLTVVQGDANSYIVDNSTPKALGKLLNPNFYK
KDKNKINTFNILNADIEDGKLLVLRGRCAGSCTLLRVIANQRESYIISVGDVTV
GNLASOMSKYRGETLYIGEDIIHPTLRGTDLTDLTKLPGRNLRPEESKRNFTK
IYELVSMGLVNOGDMVGNFVRLSGSGGKKRITITEAMVSGSSTCWCDCSTGRLD
AASAYDYINQSDTLDKTTIASFYQASDSIYNLEKVIKLDGRCIYFEGPTGLAK
QYFDLGDPCDEPKQSPDFKQVLEKNSRTTSKNSYSSSFVAORTALTORLSLT
MKHAQLEYEKQVQKSPDFKQVLEKNSRTTSKNSYSSSFVAORTALTORLSLT
WGDFTTSLRFLTILVLSIFGGIVFOOPLTDLGLFEGGAIFTSIFPNCILTOGELH
GALSRRILQKHYSALYRPSAYFVSQLLIDIPILVOVFLHSFIVFYMYFEYRADK
FFIFCFTLVGSLSSASLFRGFANFTQNLNMFVIFEFYNYGYSQTPDKMS
WFKYTIINLAYALQSLMINEFKGLDFSCIDSAIPDFHNNSTYSMSHRSYVPS
LEGSLVKGYNLMDALQINSDHRLANVVVILFWLFEYIGLNFVAYFDWTSYVTH
KVYGRKAPKLDVDEERNQIVYKATDNMKTLMKRGGLFSKNSYSYTVPVAGTNN
LLDLDIMGWIKPQOMTALMSGGAGKTTLLDVLAKRMTGTGTGSELNGKQLEIDE
RITGVQEMDVHNPGLVREALREFSAKLOBPVPLDKQYQYVHVHEMEMKHLGDA
LIGTLETVGVSVEERKRLTIGVELVAKPQILFDEPTSGLDQSSINIIFIRKLD
AGMPLVCTIHOPFSVLFHFDRILLLARGGKTVYFGDIDGKSKTLTSYFERHGVPC
ESNPAETILEATGAGHKGKTDVNNPEAWKQSSYQNVNELDLRTKEELGYIGTF
DLQVDGKQAPREFANGFLQFIEVYKRLNIYYRVDYTMGSPAQSAVSLVIGTF
YDLKNSSSDOOQRIFFMSWEAMILGVLLIYLPLWFFIQKEYKRDRTASKYYSWHA
FSL

Query Match
Best Local Similarity 5.5%; Score 60.8; DB 2; Length 129360;
Matches 415; Conservative 0; Mismatches 522; Indels 11; Gaps 3;

Qy 3 TCTCAAAATTAATAATATTAGTCAATAACAATGCATAGAAAGTTCACAAAAAATTTTG 62
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19200 TATTATATATTTTATTATATATATATTTGTTGGGGGTTGTTGTTAATTATGTTG 19141

Qy 63 TTACACAAACTTCCAAATTTTTTTTTTTTATGGAACAAGAAATACAGATAGAAAACTA 122
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19140 TTTATAAAATGGGTAGTTTTTTTTTTTTTTTTCGAAAAAATAAAATATAAATTA 19081

Qy 123 TTTTGTGTGGAAGTAGTAGTATACATATAGCAAAATTTTAAAAATATATATAAGC 182
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19080 ATAAAAATG---GGAATTTGTATTTTAAATAAAGTATAATATAAAGACACCCCT 19025

Qy 183 CTATACCGCTCAAGATGTTATCTAGTAGGTGTAATTAATATGATCGTGCATTCATCA 242
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19024 GTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTATACTATAATCAATTTAAAAATAA 18965

Qy 243 GAATGGGACACAAATCAAAACGGAATTAATAATTAACCTTTAAATAATAAAATTTG 302
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18964 AAAAAAATAAATTAATAGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 18905

Qy 303 AGTAAATGTGTTTCTGACTATTGAGGGGCAAAAAAGCAATGCCAAAGTCTACGGG 362
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18904 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 18845

Qy 363 TTTGACTGTCAGTTCGGTAATTAATCTATAACTCTGCTTTCGACCGAGCTCGTAG 422
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18844 ATTTCTAGTCGCGCTGCTGCTGCTCTCTATGAATATAAATGTTGTAATTTCTGTAAT 18785

Qy 423 GGTCTCTCTGACATTTTCTACTCTCTACCCCTACTCGTACGCCACCCCTTTCCCATAT 482
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18784 AGAATCTATTAATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 18725

Qy 483 CTTAAGGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCTCG 542
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18724 AATATTATTAATTCGGTATATACATATAAAAAAATAAATAAATAAATAAATAAATAA 18665

Qy 543 GATTCGCTGGAGCATTTTATCAAAAAATTTAGCACGAATGGGTTTATTAATTTAAAAAC 602
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18664 TTTTTTTTTT-----TTTTTAATATACAAATTTAGTTAGTTAGTTAGTTAGTTAGTT 18611

Qy 603 TCACAACTTGTATCAGATAAAATTTTCAAAACTTTTACGATGGATTCGTACGATCATC 662
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18610 TCTCCTTTTTTCCCTTTAAATGAATGAAAAAATAAATAAATAAATAAATAAATAAATAA 18551

Qy 663 TAATGACTTTTTTTTTTCTACCAGGTGATGAAAGTTATAGTACTATTATAGCCAGACA 722
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      18550 AAAAAAAAAAATAAAAAAAAAAAAAAAAAAATAATAATACTTGAATGGGATAATCT 18491
QY      723 ATT-GATTATAGATATCCATTAATCCATGATATTTATGATATATAAATAGCTGTTAACT 781
Db      18490 CTTCAATAATAGTAATCACTACTATAATAAAGTTTATAAATAATATTTATTTGTTAACTACT 18431
QY      782 ATTTCAGCATCGCAGCTTCTCAGACTTTTGTGTTTAAATTTAGAGCTTTAAATAAATAA 841
Db      18430 CTTTAAAAACAATAAGAAATAAACAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 18371
QY      842 GTATTAAAGGAGCATAACGAGGACACAAAAAGTAATGAACGAGGAGAAACAAAGCCATG 901
Db      18370 AATAAAACCCATTAATAATTTTAAAGAGATAAATAATTTATTTGTTATTTAATAATAAT 18311
QY      902 AAGCTCATTTGTTAGTTTAAAGCTTAATAAGACAGATTTTAAATTTT 949
Db      18310 TTATTTTTATTTTATTTTATTTTGTGTTTATTTTATTTTATTTTATTTT 18263

RESULT 6
TSP418778/c
LOCUS      TSP418778      3364 bp      RNA      linear      VRL 02-MAY-2002
DEFINITION Tomato spotted wilt virus NSs gene and N gene, genomic RNA, isolate 10HK96.
ACCESSION AJ418778
VERSION    1
KEYWORDS   N gene; N protein; NSs gene; NSs protein.
SOURCE     Tomato spotted wilt virus.
ORGANISM   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
REFERENCE 1
AUTHORS    Heinze,C., Willingmann,P. and Adam,G.
TITLE      Short intergenic regions of the S RNAs of Tomato spotted wilt
JOURNAL    tospovirus -not a species characteristic
AUTHORS    Unpublished
TITLE      2 (bases 1 to 3364)
JOURNAL    Direct Submission
AUTHORS    Adam,G.
TITLE      Submitted (05-NOV-2001) Adam G., Pflanzenschutz, Institut fuer
JOURNAL    Angewandte Botanik, Ohnhorststrasse 18, 22609 Hamburg, GERMANY
FEATURES   source
            1..3364
            /organism="Tomato spotted wilt virus"
            /viral
            /isolate="10HK96"
            /specific_host="virginia tobacco"
            /db_xref="taxon:11613"
            /country="Bulgaria"
            /gene="NSs"
            /gene="NSs"
            /codon_start=1
            /product="NSs protein"
            /protein_id="CAD11445.1"
            /db_xref="GI:20429082"
            /translation="MSSSVYESIIQTRASVWGVSAGKAVVDSYWIHELGTCSLLVQT
            QLYSDRSKSGFYTKVKNLPCEEEILSQRYPIEDDFIDFSINIDDSVIALSVCS
            NTVNTGKHQGLHLVSPDLQHSIGTSMNRSDIKRQLOEKDIPNDRYIEAANKG
            SLSCVKEHYKIETCYNOALGVNVLSPNRNHWLYSKFSENOVESNNRTVNSLAV
            RSLTSAENNPNSQAFVKASTDSHKFLSLWLPKLVROVSTOKLEFKAGDENKT
            FYLSVICPNHNSVETALNISVICKHQLPIKCKAPSELSWFSOLKEPYNIVHDPSY
            PORVHALLETHTSFQVLCNNLQEDVIIITLNNVLPKLDLGERLTNYSEDICR
            IYFKSLTECLPSNTQMSYLDISQIPSWKIDFARGGKIKISQPVSVAKSLKLDLSG
            IKKGSKISDGYASGSK"
            complement(2436..3212)
            /gene="N"
            /complement(2436..3212)
            /gene="N"
            /function="structural protein"
            /codon_start=1
            /product="N protein"
            /protein_id="CAD11446.1"

gene
CDS

gene
CDS

```

```

/db_xref="GI:20429083"
/translation="MSKVLTRESIVALLTQKDLFEEDQNLVAFNFKTFCLENDQ
IKRMSVISCLTFPKNRQSIMKVIKQSDFTFGKITIKKTSRIGATQMTFRRLDSMRV
RIVEETGNSENLNTIKSKIASHPLOIAYGLPLDDAKSVRLAIMGGSLPLIASVDSFE
MISVLAIQDANTDKLIDPKKYDTKEALGKVCTVLKSKAFEMNEDQVKGREYAAI
LSSNPNAGSGINMGHYSETLKNKFFYEMFGVKKQAKLTELA"
BASE COUNT 1002 a 563 c 489 g 1310 t
ORIGIN

Query Match      5.5%; Score 60.6; DB 14; Length 3364;
Best Local Similarity 48.2%; Pred. No. 0.38;
Matches 171; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY      1 AATCTACAATTTAATAATTTAGTCAATAACAATGCATAGAAAGTTCCAAAAAATTT 60
Db      2100 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAAT 2041
QY      61 TGTTAACAGAACTTCCAAATTTTTTTTTTTTATGGAACAGAAATAACAGATAGAAAAC 120
Db      2040 AAAAAATAAAAGTAAAAATAAAGAAATAAAAAATAAAAAATAAAAAATAAAAA 1981
QY      121 TATTTTTGTGGAATGGAAGTAGTAAATATACATTAAGCAAAATTTTAAAAAATATATAA 180
Db      1980 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAA 1921
QY      181 GCCTATACGCGCTCAAAAGTATGTTATCTAGTAGTGTAAATTAATAATGTCATGCGGATT 240
Db      1920 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1861
QY      241 CAGATTGGGACACAAATGAACACGGAATTAATAATTTAACTTTAAAAATAAAAAAT 300
Db      1860 TAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAA 1801
QY      301 TGAGTAAATGTGTTTTCTGACTATTGAGGGGCAAAAAAGCAATGCCAAAAGT 355
Db      1800 ATAAATAAATAAATAAAAAACAAAAACAAAAACAAAAACAAAAACCAAAAT 1746

RESULT 7
AC020941/c
LOCUS      AC020941      171317 bp      DNA      linear      PRI 01-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-231219, complete sequence.
ACCESSION AC020941
VERSION    AC020941.5
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171317)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Unpublished
AUTHORS     2 (bases 1 to 171317)
TITLE      DOE Joint Genome Institute.
JOURNAL    DOE Joint Genome Institute.
AUTHORS     Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
TITLE      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL    3 (bases 1 to 171317)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS     Drive, Walnut Creek, CA 94598, USA
TITLE      On Jun 1, 2001 this sequence version replaced gi:12830142.
JOURNAL    Draft Sequence Produced by DOE Joint Genome Institute
COMMENT     www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.7% of Sequence;
            Estimated Total Number of Errors is 2.7.
            Location/Qualifiers
            1..171317
            /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="CTD-21219"
BASE COUNT 50319 a 31872 c 34147 g 54979 t
ORIGIN
Query Match 5.5%; Score 60.6; DB 9; Length 171317;
Best Local Similarity 49.8%; Pred. No. 0.17;
Matches 153; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 2 ATCTACAAATATATATATATATAGTCAATAAAGTCAATAGAGTTCACAAAAAATTTT 61
Db 22530 ATATAAATATATATAAATATATATAATATAAATATATAAATATAAATAATAT 22471
QY 62 GTTACAGAAACTTCCAAATTTTATGGAACAAGAAATACAGATAGAAAAC 121
Db 22470 ATATATATATATAAATATATATATATATATATATATATAAATATATAT 22411
QY 122 ATTTGTTGGAATGGAAGTAGTATATATACATTAAGCAATTTTAAAAAATATATAAG 181
Db 22410 TATATATATATATAAATATATAAATATATATATATATATAAATATAAATATA 22351
QY 182 CCTACACGGCTCAAGATGATCTAGTAGTCTAATTAATATATATATATATATAT 241
Db 22350 TATATATATATAAATATATAAATATATATATATATATATATATAAATATATA 22291
QY 242 AGAATGGGACAAATGAAACGGAATTAATAATTAATTAATTAATAAATAAATAATTT 301
Db 22290 AAATATATATATATATATATATATATATATATATATATATATATATATATAT 22231
QY 302 GAGTAAA 308
Db 22230 ATATAAA 22224

```

```

RESULT 8
AC094291
LOCUS
DEFINITION
Rattus norvegicus clone CH230-3H24, *** SEQUENCING IN PROGRESS ***,
60 unordered pieces.
AC094291
AC094291.3 GI:21716642
VERSION
HTG: HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
AUTHORS
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbata, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowles, S., Brice, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyne, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escott, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssohn, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,

```

```

Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.D., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, K.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 233269)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233269)
Worley, K.C.
Direct Submission
Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941008.
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Center project name: GAIL
Center clone name: CH230-3H24
-----
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175720 bases at least Q40
Consensus quality: 180089 bases at least Q30
Consensus quality: 183792 bases at least Q20
-----

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1068: contig of 1068 bp in length
* 1069 1168: gap of unknown length
* 1169 2363: contig of 1195 bp in length
* 2364 2463: gap of unknown length
* 2464 3682: contig of 1219 bp in length
* 3683 3782: gap of unknown length
* 3783 4897: contig of 1115 bp in length
* 4898 4997: gap of unknown length
* 4998 6251: contig of 1253 bp in length
* 6251 6351: gap of unknown length
* 6351 7467: contig of 1117 bp in length
* 7468 8742: contig of 1175 bp in length
* 8743 8842: gap of unknown length
* 8843 10039: contig of 1197 bp in length

```

*	10040	10139:	gap of	unknown	length
*	10140	11793:	contig of	1654	bp in length
*	11794	11893:	gap of	unknown	length
*	11894	13711:	contig of	1478	bp in length
*	13372	13471:	gap of	unknown	length
*	13472	14864:	contig of	1393	bp in length
*	14865	14964:	gap of	unknown	length
*	14965	16460:	contig of	1496	bp in length
*	16461	16560:	gap of	unknown	length
*	16561	17663:	contig of	1103	bp in length
*	17664	17763:	gap of	unknown	length
*	17764	19281:	contig of	1518	bp in length
*	19282	19381:	gap of	unknown	length
*	19382	21112:	contig of	1731	bp in length
*	21113	21212:	gap of	unknown	length
*	21213	22833:	contig of	1621	bp in length
*	22834	22933:	gap of	unknown	length
*	22934	24214:	contig of	1281	bp in length
*	24315	24314:	gap of	unknown	length
*	24315	25527:	contig of	1213	bp in length
*	25528	25627:	gap of	unknown	length
*	25628	27220:	contig of	1593	bp in length
*	27221	27320:	gap of	unknown	length
*	27321	28926:	contig of	1606	bp in length
*	28927	29026:	gap of	unknown	length
*	29027	30363:	contig of	1337	bp in length
*	30364	30463:	gap of	unknown	length
*	30464	31665:	contig of	1202	bp in length
*	31666	31765:	gap of	unknown	length
*	31766	33449:	contig of	1584	bp in length
*	33450	33449:	gap of	unknown	length
*	33450	36317:	contig of	2868	bp in length
*	36318	36417:	gap of	unknown	length
*	36418	37778:	contig of	1361	bp in length
*	37779	37878:	gap of	unknown	length
*	37879	39851:	contig of	1973	bp in length
*	39852	39951:	gap of	unknown	length
*	39952	42802:	contig of	2851	bp in length
*	42803	42902:	gap of	unknown	length
*	42903	45682:	contig of	2780	bp in length
*	45683	45782:	gap of	unknown	length
*	45783	49268:	contig of	3486	bp in length
*	49269	49368:	gap of	unknown	length
*	49369	51428:	contig of	2060	bp in length
*	51429	51528:	gap of	unknown	length
*	51529	54571:	contig of	3043	bp in length
*	54572	54671:	gap of	unknown	length
*	54672	57546:	contig of	2875	bp in length
*	57547	57646:	gap of	unknown	length
*	57647	60740:	contig of	3094	bp in length
*	60741	60840:	gap of	unknown	length
*	60841	63916:	contig of	3076	bp in length
*	63917	64016:	gap of	unknown	length
*	64017	67000:	contig of	2984	bp in length
*	67001	67100:	gap of	unknown	length
*	67101	70219:	contig of	3119	bp in length
*	70220	70319:	gap of	unknown	length
*	70320	72950:	contig of	2631	bp in length
*	72951	73050:	gap of	unknown	length
*	73051	76802:	contig of	3752	bp in length
*	76803	76902:	gap of	unknown	length
*	76903	79990:	contig of	3088	bp in length
*	79991	80090:	gap of	unknown	length
*	80091	82745:	contig of	2655	bp in length
*	82746	82845:	gap of	unknown	length
*	82846	87725:	contig of	4880	bp in length
*	87726	87825:	gap of	unknown	length
*	87826	91178:	contig of	3353	bp in length
*	91179	91278:	gap of	unknown	length
*	91279	94533:	contig of	3255	bp in length
*	94534	94633:	gap of	unknown	length
*	94634	99871:	contig of	5238	bp in length
*	99872	99971:	gap of	unknown	length

		5.48; Score 60; DB 2; Length 233269;	
Query Match	Best Local Similarity	51.3%; Pred. No. 0.2;	
Matches 138; Conservative	0; Mismatches 131; Indels	0; Gaps 0;	
Qy	6	ACAAATTATTAATATTTAGTCAATAACAATGCATAGAAAGTCCCAAAAAAATTTTGGTTA	65
Db	1391	AAAAAAATTTTCTGTTTTTTTTTAAAAAGTTAAAAAAAATTTTAAAAAATTTTTTTTTTTT	1450
Qy	66	ACAGAAACTTCCAATTTTTTTTTTTTTTTATCGACAAGAAATACAGATAGAAAACATATTT	125
Db	1451	TTTTTAAAAAAATATTTTTTTTTTTTTTTTGGAAAAAANAANAANAANAANAATATATTT	1510
Qy	126	TGTTGTGGAATGGAAGTAGTAATATACATTTAAGCAAAATTTTAAAAAATTTATATAAGCCTA	185
Db	1511	TTTTTTAAAAAAAATAAAAAAATTTTTTTTGTGGAAAAATTTTAAAAAAAATTTT	1570
Qy	186	TACGCGCTCAAGTAGTCTTATCATAGTAGGTGTAATTAATAATGCATGGTCCGATTCAGAA	245
Db	1571	TTTTTATAAAAAAAATAATTTTTTTTTTTTTTTGCAAAAAAANAATATTTTTTTTTTTT	1630
Qy	246	TTGGACAAACAAATGAAACGGAATTAATA	274
Db	1631	TTTANAAAAAAANAANAANAANAANAATAAAA	1659

RESULT 9	AL732314/c	LOCUS	AL732314	224635 bp	DNA linear	HTG 17-AUG-2002
DEFINITION			Homo sapiens chromosome X clone RP13-465B17, *** SEQUENCING IN PROGRESS ***, 8 unordered pieces.			

ACCESSION AL732314
VERSION AL732314.10 GI:22415930
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.

ORGANISM Homo sapiens
TAXONOMY Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 224635)

AUTHORS	Whitehead, S.
TITLE	Direct Submission
JOURNAL	Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hamquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On Aug 21, 2002 this sequence version replaced qi:22204483.

ON MAY 22, 2002, THE SEQUENCING VERSION 3.0.0 WAS RELEASED.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
-----
Center project name: BB465B17
-----
Summary Statistics
-----
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads

```

```

Consensus quality: 222703 bases at least Q40
Consensus quality: 223253 bases at least Q30
Consensus quality: 223562 bases at least Q20
Insert size: 223935; sum-of-contigs
Insert size: 183401; 20.9% error; agarose-fp
Quality coverage: 6.54x in Q20 bases; sum-of-contigs Quality
coverage: 8.94x in Q20 bases; agarose-fp
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
* 3853 3952: gap of 100 bp
* 3953 36565: contig of 32613 bp in length
* 36566 36665: gap of 100 bp
* 36666 70753: contig of 34088 bp in length
* 70754 70853: gap of 100 bp
* 70854 141740: contig of 70887 bp in length
* 141741 141840: gap of 100 bp
* 141841 180105: contig of 38265 bp in length
* 180106 180205: gap of 100 bp
* 180206 187608: contig of 7403 bp in length
* 187609 187708: gap of 100 bp
* 187709 191079: contig of 3371 bp in length
* 191080 191179: gap of 100 bp
* 191180 224635: contig of 33456 bp in length.
FEATURES
      Location/Qualifiers
      source
        1..224635
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="X"
          /clone_lib="RPCI-13.2"
          1..3852
            /note="assembly_fragment:02748
            fragment_chain:1"
            3953..36565
              /note="assembly_fragment:04038
              fragment_chain:1"
              36666..70753
                /note="assembly_fragment:01967
                fragment_chain:1"
                70854..141740
                  /note="assembly_fragment:03650
                  fragment_chain:1"
                  141841..180105
                    /note="assembly_fragment:03847
                    fragment_chain:1"
                    180206..187608
                      /note="assembly_fragment:01716"
                      187709..191079
                        /note="assembly_fragment:02259"
                        191180..224635
                          /note="assembly_fragment:04008"
                          BASE COUNT 56417 a 52985 c 53054 g 61479 t 700 others
ORIGIN
Query Match 5.4%; Score 59.4; DB 2; Length 224635;
Best Local Similarity 48.4%; Pred. No. 0.26;
Matches 165; Conservative 0; Mismatches 176; Indels 0; Gaps 0;
QY 6 ACAATTTATTTAGTCTCAATACATGCATAGAAAGTCCAAAAAAATTTTGTA 65
Db 197253 AGAATAAATTTACTTAAATAAATTTTCATATTAATTTGTAATAATAAATTTAT 197194
QY 66 ACAGAAACTTCCAAATTTTTTTTTTTTTTTATGCAACAAGAAATACAGATAGAAACTATTT 125
Db 197193 ATTTAAATAATAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAT 197134

```


Query Match	5.4%	Score 59	DB 3	Length 253305
Best Local Similarity	47.0%	Pred. No. 0.3		
Matches 182	Conservative	0	Mismatches 205	Indels 0
QY	3	TCTACAATATTAAATATTAGTCAATACAAATGCATAGAAAGTTCCTCAAAAAAATTTTG	62	
Db	156472	TATAAANATTTTAATTTTATTAATAAATAAATACTTTTAAATAACAAAAAATAAAAA	156531	
QY	63	TTACAGAAACTTCCAAATTTTTTTTTTTTATGGAACAAGAAATACAGATAGAAAACTA	122	
Db	156532	AAAAAATAAAGAACATATAATTTATTTAAATGTAACTGAAGATAAAATAAATAAAA	156591	
QY	123	TTTTTGTGTGGATGGAAGTAGTATATACATTAAGCAAAATTTAAAAAATATATAAGC	182	
Db	156592	TAAATGTAANATTTAAATATTTTTCATAAATGTATATATATATATATAAATAAT	156651	
QY	183	CTATACGGCGTCAAAAGTAGTATCTAGTAGGTCTAATTAATTAATCGATGGTGCATCA	242	
Db	156652	ATATATAATAGAAAAATAAATATAAATTTTAATTATACATTTAAAAATAAATTTA	156711	
QY	243	GAATGGGACACAAATGAAAACGGAAATTAANAATTAACTTTAAAAATAAATAAATTTG	302	
Db	156712	TATTAATAGTAAAAATATATATATATATATATATATATATATATTAATATATATTTT	156771	
QY	303	AGTAATGTGTTTCTGACTATTGAGGGGCAAAAAAGACAATGCCAAAAGTCTACGGG	362	
Db	156772	ACATATATATATATAATATATTTAAATTAATAATTTACGAAATTTGAATATATACATA	156831	
QY	363	TTTGACTCTCAGTTCGGTAAATCT	389	
Db	156832	ATTATATATATATATATAAATAAT	156858	

RESULT 11	AC114238	165669 bp	DNA	linear	HTG 13-JUL-2002
LOCUS	AC114238	Rattus norvegicus clone CH230-336N7, ***	SEQUENCING IN PROGRESS		
DEFINITION	***, 49 unordered pieces.				
ACCESSION	AC114238				
VERSION	AC114238.6	GI:21738343			
KEYWORDS	HTG; HTGS, PHASE1.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
	Rattus.				
REFERENCE	1 (bases 1 to 165669)				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhadj,C., Escotto,M., Falls,T., Ferraguto,D., Flagn,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsli,E., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,				

Masey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwunodu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Plickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Soneake, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 165669)
Worley, K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165669)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:20467600.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GQXB
Center clone name: CH230-336N7
----- Summary Statistics -----
Sequencing vector: plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 128160 bases at least Q40
Consensus quality: 133217 bases at least Q30
Consensus quality: 136246 bases at least Q20

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1074: contig of 1074 bp in length
* 1075
* 1174: gap of unknown length
* 1175
* 2179: contig of 1005 bp in length
* 2180
* 2279: gap of unknown length
* 2280
* 3300: contig of 1021 bp in length
* 3301
* 3400: gap of unknown length
* 3401
* 4592: contig of 1192 bp in length
* 4593
* 4692: gap of unknown length
* 4693
* 6219: contig of 1527 bp in length
* 6220
* 6319: gap of unknown length
* 7439: contig of 1120 bp in length
* 7539: gap of unknown length
* 7540
* 9318: contig of 1779 bp in length
* 9319
* 9418: gap of unknown length
* 9419
* 10930: contig of 1512 bp in length
* 10931
* 11030: gap of unknown length

```


*	11031	13469:	contig	of	2439	bp	in	length
*	13470	13569:	gap	of	unknown	length		
*	13570	14675:	contig	of	1106	bp	in	length
*	14676	14775:	gap	of	unknown	length		
*	14776	16102:	contig	of	1327	bp	in	length
*	16103	16202:	gap	of	unknown	length		
*	16203	18320:	contig	of	2118	bp	in	length
*	18321	18420:	gap	of	unknown	length		
*	18421	20658:	contig	of	2238	bp	in	length
*	20659	20758:	gap	of	unknown	length		
*	20759	22073:	contig	of	1315	bp	in	length
*	22074	22173:	gap	of	unknown	length		
*	22174	24783:	contig	of	2610	bp	in	length
*	24784	24883:	gap	of	unknown	length		
*	24884	27460:	contig	of	2577	bp	in	length
*	27461	27560:	gap	of	unknown	length		
*	27561	29143:	contig	of	1583	bp	in	length
*	29144	29243:	gap	of	unknown	length		
*	29244	31381:	contig	of	2138	bp	in	length
*	31382	31481:	gap	of	unknown	length		
*	31482	33731:	contig	of	2250	bp	in	length
*	33732	33831:	gap	of	unknown	length		
*	33832	36867:	contig	of	3036	bp	in	length
*	36868	36967:	gap	of	unknown	length		
*	36968	39884:	contig	of	2917	bp	in	length
*	39885	39984:	gap	of	unknown	length		
*	39985	42654:	contig	of	2670	bp	in	length
*	42655	42754:	gap	of	unknown	length		
*	42755	44617:	contig	of	1863	bp	in	length
*	44618	44717:	gap	of	unknown	length		
*	44718	46244:	contig	of	1527	bp	in	length
*	46245	46344:	gap	of	unknown	length		
*	46345	49733:	contig	of	3389	bp	in	length
*	49734	49833:	gap	of	unknown	length		
*	49834	52093:	contig	of	2260	bp	in	length
*	52094	52193:	gap	of	unknown	length		
*	52194	54926:	contig	of	2733	bp	in	length
*	54927	55026:	gap	of	unknown	length		
*	55027	58424:	contig	of	3398	bp	in	length
*	58425	58524:	gap	of	unknown	length		
*	58525	61513:	contig	of	2989	bp	in	length
*	61514	61613:	gap	of	unknown	length		
*	61614	63685:	contig	of	2072	bp	in	length
*	63686	63785:	gap	of	unknown	length		
*	63786	66780:	contig	of	2995	bp	in	length
*	66781	66880:	gap	of	unknown	length		
*	66881	70124:	contig	of	3244	bp	in	length
*	70125	70224:	gap	of	unknown	length		
*	70225	74001:	contig	of	3777	bp	in	length
*	74002	74101:	gap	of	unknown	length		
*	74102	77381:	contig	of	3280	bp	in	length
*	77382	77481:	gap	of	unknown	length		
*	77482	80568:	contig	of	3088	bp	in	length
*	80570	80669:	gap	of	unknown	length		
*	80670	83635:	contig	of	2966	bp	in	length
*	83636	83735:	gap	of	unknown	length		
*	83736	87660:	contig	of	3925	bp	in	length
*	87661	87760:	gap	of	unknown	length		
*	87761	90614:	contig	of	2834	bp	in	length
*	90615	90714:	gap	of	unknown	length		
*	90715	95460:	contig	of	4746	bp	in	length
*	95461	95560:	gap	of	unknown	length		
*	95561	100030:	contig	of	4470	bp	in	length
*	100031	100130:	gap	of	unknown	length		

[illegible]

RESULTS 12

AC007182 192389 bp DNA linear PRI 25-AUG-1999
 Homo sapiens chromosome 14 clone BAC 507E23 map 14q24.3, complete
 sequence.
 AC007182
 AC007182.3 GI:5708446
 HTG.
 Homo sapiens.
 Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 192389)
 Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
 Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
 Madan,A., Ratcliffe,A., Shafer,T. and Hood,L.
 Submitted (30-MAR-1999) Multimegabase Sequencing Center, University
 of Washington, PO BOX 357730, Seattle, WA 98195, USA
 2 (bases 1 to 192389)
 Young,J., Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M.,
 Dickhoff,R., James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S.,
 Ratcliffe,A., Shafer,T. and Hood,L.
 Direct Submission
 Submitted (30-MAR-1999) Multimegabase Sequencing Center, University
 of Washington, PO BOX 357730, Seattle, WA 98195, USA
 3 (bases 1 to 192389)
 Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
 Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
 Madan,A., Ratcliffe,A., Shafer,T. and Hood,L.
 Direct Submission
 Submitted (25-AUG-1999) Multimegabase Sequencing Center, University
 of Washington, PO BOX 357730, Seattle, WA 98195, USA


```

/feature="ORF_ID:dd_01651"
/codon_start=1
/product="Sequence 3 from Patent WO0075321. 6/101"
/protein_id="AA33154.1"
/db_xref="GI:20376565"
/translation="MNOVAESINKSKSTADYTPKGYHITELSDSYLDHFSGN
AKLFHPNPAVAVLYEQALAEHGSATITSGALVTRSGVTKGRSPKDKRIVKPESSODDI
WGPVNIMDDLSIFMTNREARADYLNTOEKIYVIDYAGWDPKYRKIVKVICARAYHA
LFHNNMLIRPANREELNREGEPTDIYNAGOPFANRYTKMGSSSSSIALIDFARKEMVI
LGTOYAGEMKGLITIMYLMKPMGVPLHSSCQARNNGDITLFFGLSGTGKTTLSA
DINRELIGDDEHWDTGCFNIEGGCYAKCIDLSREKEPEIFDAIFGAVLENNVYNE
YSRKYDYNDSYENTRCAPLEHIPNAKFPAYAGHPKNIIMLTCDAFGLIPVSRLO
AQVMHIFLQGTAVAGTEVTEPTAFSSCYGEPTVWHPTKYAEMLASOLHKHS
ARAWLINTGWSHGSGRIKLAITRAIDAIHSGELEKIPTTKMDVFGQVQNSCP
GYPSLEIMPINGWADKEKYSVTHMKLAKLFIENFKFKODKASPELVAAGPIIPQ"
join(28969..29084,29176..29317,29453..30610)
/note="ORF_ID:dd_00794"
/codon_start=1

```

```

Query Match 5.3%; Score 58; DB 2; Length 115489;
Best Local Similarity 52.0%; Pred. No. 0.52;
Matches 155; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 1 AATCTACAAATTAATAATTTAGTCAATACAAATGCATAGAAATTCACAAAAAATTT 60
   || || || || || || || || || || || || || || || || || || || || ||
Db 66587 AAAAAAATAATTAATAATAAATAATTAATAATAGAAATTCCTTCAAAAAA 66528

QY 61 TCTTACAGAACTTCCAAATTTTTTTTTTATGGAACAAGAAATACACATAGAAAC 120
   || || || || || || || || || || || || || || || || || || || || ||
Db 66527 AAAAAAATAATTCATATAATGATGTTTTTTTAAAGGCAAAAAAATAA--AAAAAAT 66471

QY 121 TATTTTGTGTGAATGGAATGAGTAAATATACATTAAGCAATTTTAAAAATATATAA 180
   || || || || || || || || || || || || || || || || || || || || ||
Db 66470 TTTATATTGATATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 66411

QY 181 GCCTATAGCGCTCAAGATGTTATCTATGATGAGTAAATTAATAATCCATGTCGAT 240
   || || || || || || || || || || || || || || || || || || || || ||
Db 66410 TAGGAACACATTCAAATTTTATTTTATTTGCTCCACATAGATAAATAAGAAAT 66351

QY 241 CAGATTGGGACACATGAACGGAATTAATAATATTAATTTTAAATAAATAAATAA 298
   || || || || || || || || || || || || || || || || || || || || ||
Db 66350 TCCACTATTACAAATTTAAATAATAATAATAATAAATAAATAAATAAATAAATAA 66293

```

```

RESULT 14
AC115575/c
LOCUS
DEFINITION
  Dictyostelium discoideum chromosome 2 map 1180800-1208089 strain
  AX4,*** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION
  AC115575
VERSION
  HTG; HTGS.PHASE2.
KEYWORDS
  Dictyostelium discoideum.
SOURCE
  Dictyostelium discoideum
  ORGANISM
    Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
  1 (bases 1 to 27291)
  Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
  Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
  Tünggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
  Noegel,A.A.
  Sequence and Analysis of Chromosome 2 of Dictyostelium
  Unpublished
  The Dictyostelium Genome Sequencing Consortium
  2 (bases 1 to 27291)
  Baumgart,C.
  Direct Submission
  Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
  Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
  CDS predictions from GenBank may contain errors. Further information
  is available from IMB Jena, Department of Genome Analysis
  (http://genome.imb-jena.de/dictyostelium/)
  and the University Cologne, Institute for Biochemistry I
  (http://www.uni-koeln.de/dictyostelium/project.shtml)

```

Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source

```

1..27291
Location/Qualifiers
  /organism="Dictyostelium discoideum"
  /strain="AX4"
  /db_xref="taxon:44689"
  /chromosome="2"
  /map="1180800-1208089"
  complement(join(163..380,1237..2062,2157..2468))
  /note="ORF_ID:dd_00011"
  /codon_start=1
  /product="Hypothetical protein yfew precursor"
  /protein_id="AAL92187.1"
  /db_xref="GI:19569868"
  /translation="MKYINILLFVIIGCVVSLKSPIKSIDVRLNNOQVYPSD
  ETWSEVDTFLEEAIENGTFPCVAVLGNOKGVLYASAKGSFTYIGTPIINDEKVPME
  LSTLFDMASCSKVTAFTSVAQVORGEINLDSPTDFLEGEYKSGKSGISVRCNLL
  HNSGFYDPNPFWNTAEFNCPEATEEYPPMEFTQCQOIYQSIILNSLMPNIGSTYYIS
  DLNEMTLMYVGNLAKOFNITADQLTPGCDNGAPENAOQYEAAYVRHYFVNALNLV
  TNYLPGYLAQCAPENTDVMHTTIQGVYSDGNAYALGDISGHAGVSNVEMWIF
  MYSLMFAEANSQYLNSTVQYFTTEYNHTQSSRALGNWNTNDPSVFDGEGWGLACGTLSA
  KTWMLGYTGTMLCGDPERELIVILLNRYVDPNSNIKIENVRKPFSTLVQVYDSII
  N"
  complement(join(2838..3347,3546..3650))
  /note="ORF_ID:dd_00013"
  /codon_start=1
  /product="HSPC305 (FRAGMENT). 10/100"
  /protein_id="AAL92188.1"
  /db_xref="GI:19569869"
  /translation="MTDNNIFGNTKRFNDSIELINKLETPRQKILTRVNERLTFESE
  ENLQNLKFTLSLEFKILECCSFIFEQTAHYSLSPNNLVNQLKTKMLNDDKTSFQS
  VWBDSHEVLNLFLEKTOSIAPLOLNEIGRHLHYOMSSSTAIKRNASIMELNENNNNN
  NNNNNNNNNINNDKMLLEFNKEQLLEFYNKLESIQEKLDLIA"
  join(3837..3886,4358..5195)
  /note="ORF_ID:dd_00015"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="AAL92189.1"
  /db_xref="GI:19569870"
  /translation="MGIFRLSPTIIGIINCNSGKICLNKQNSCNQYSIEKINNKTQ
  QQLLNKLKINKLVNKSPLSKINHFINKESNNIIVGSIKNNINGNDLNVIRVY
  KQLPLGNKEITDKYTLNNNPFCLNNTNGGKFCYQLISTLVNYYNNNYLSKI
  IYFQBNVKYFDNNWNLAYKSVIQDSKLIAGLTINNMMVSVNAYINIPDIEKCI
  PPKSKCNQSQIITNSRDFNRLCTSTCTAFNGTWTWVIPIYPTCPKGYLTFGTG
  NNGGKLEFNCDAFLTDIV"
  join(6320..6707,6802..7274)
  /note="ORF_ID:dd_00017"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="AAL92190.1"
  /db_xref="GI:19569871"
  /translation="MKLFCISAILIFIFISGESTLIRPNFYRIDNENINCIRAPC
  POYSAQKVTNENSTAILSFYIDGYNKSYLSEARSLIFGSIQPNPPEAKDL
  KVVRYKSLPLGNPASADTKYIFDNGVRCTVTPCASTDAVLNIRTKQVSDILMP
  YSKNVGYFDSAWGSKTVSDDYGLIGGTFSGKTIISNAFIHLDPDIPKCDLPL
  LKCPGGSIMTYSRNENRCLASPTCTQSFCTLGMPLCPAGYRLDQFRATEANGCPKY
  CDPTFVCKTH"
  join(7879..8275,8373..8845)
  /note="ORF_ID:dd_00018"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="AAL92191.1"
  /db_xref="GI:19569872"
  /translation="MKYKILLTFTTFFITFVSGSSILKANFYKVDWENYOCFV
  PPCPKYTIQKLTNEKSKQILDFIPNGLNKTTLIGDESKTLIVLGTQPSKPFNNA
  TDFLVTRYKSLPLGNKOTSDKYMFEGDNGTRCKTFCPCPNIVAAALNVHSQELINTI
  SQPYEKVNGFLDSWLSKSNIRSDFGILGIGTKNGVISISNSFIPLPDIPKCPKL
  PLLKCVGNSMTYSRDENRCLVSPRCFVGCTLSILPCKNKGFLVSPFSTELNGCPK

```


Search completed: June 11, 2003, 01:49:57
Job time : 2938.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 21:59:33 ; Search time 252.042 Seconds

(without alignments)
9837.455 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatatacaattattaatat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	6888	21	AAA59599
2	56.6	5.1	6107	24	ABL70390
3	56.6	5.1	6107	24	AAS61342
4	56.6	5.1	6107	24	ABK31431
5	54.2	4.9	875	21	AAA01920
6	54.2	4.9	6113	24	ABL32803
7	53.8	4.9	37973	24	ABL34197
8	53.8	4.9	83391	24	ABQ67094
9	53.4	4.9	17131	24	ABL33053

10	53.2	4.8	8170	24	ABK28257	DNA transcription
11	53	4.8	54108	24	ABK22782	Human high bone ma
12	53	4.8	57273	24	ABK22784	Human high bone ma
13	53	4.8	66933	22	ABAB2625	Human HBM gene reg
14	53	4.8	72049	22	ABAB2623	Human HBM gene reg
15	52.6	4.8	16228	24	ABL70459	Chemically treated
16	52.6	4.8	16228	24	AAS61424	Human gene regulat
17	52.4	4.8	4501	24	ABK33968	Human DNA for stag
18	52.4	4.8	5940	21	AAA70105	Plasmodium falcipa
19	52.4	4.7	7746	24	ABK40047	Human chemically p
20	52.2	4.7	7746	24	ABL33856	Human immune syste
21	52.2	4.7	78925	21	ABK33856	Human FN gene. Ho
22	52	4.7	5309	22	AAS46527	Tumour suppressor
23	52	4.7	5309	24	ABK40039	Human chemically p
24	52	4.7	5309	24	ABL33736	Human immune syste
25	52	4.7	6636	24	ABN80023	Human chemically m
26	52	4.7	7459	24	ABK31382	Signal transductio
27	51.8	4.7	641	24	ABQ56694	Human colon cancer
28	51.6	4.7	47108	24	ABK31511	Signal transductio
29	51.4	4.7	4041	21	AAA70170	Plasmodium falcipa
30	51.2	4.7	3738	21	AAA70178	Plasmodium falcipa
31	51.2	4.7	5690	22	AAS45368	Chemically pretrea
32	51.2	4.7	5690	24	ABK28205	DNA transcription
33	51.2	4.7	23695	24	ABQ66981	Human angiogenesis
34	51	4.6	10286	22	AAS45309	Chemically pretrea
35	51	4.6	10286	24	ABK28148	DNA transcription
36	50.8	4.6	5880	22	AAS46331	Tumour suppressor
37	50.8	4.6	5880	24	ABK28177	DNA transcription
38	50.8	4.6	6254	24	ABL33621	Human immune syste
39	50.6	4.6	12405	22	AAS45330	Chemically pretrea
40	50.6	4.6	12405	24	AAS61143	Human gene regulat
41	50.6	4.6	12405	24	ABK28169	DNA transcription
42	50.4	4.6	5689	22	AAS45384	Chemically pretrea
43	50.4	4.6	5689	22	AAS46426	Tumour suppressor
44	50.4	4.6	5689	24	ABK28226	DNA transcription
45	50.4	4.6	8085	22	AAS46479	Tumour suppressor

ALIGNMENTS

RESULT 1

AAA59599

ID AAA59599 standard; DNA; 6888 BP.

XX AC AAA59599;

XX AC AAA59599;

XX DT 14-NOV-2000 (first entry)

XX DE DNA encoding a cytochrome P450 enzyme designated DWF4.

XX DE DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;

XX KW plant phenotype; cell elongation; ss.

XX OS Arabidopsis sp.

XX FH Key

XX FT promoter

XX FT TATA_signal

XX FT CDS

XX FT exon

XX FT intron

XX FT exon

XX FT intron

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

DT 01-JUL-2002 (first entry)
XX Chemically treated cell signalling DNA sequence complementary to#140.
XX Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX Unidentified.
XX WO200202807-A2.
XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-EP07471.
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-154758/20.
DR Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signalling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signalling
XX Claim 1: SEQ ID NO 280; 24ppt+sequence listing; English.
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL/0111-ABL/0626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;
Query Match 5.1%; Score 56.6; DB 24; Length 6107;
Best Local Similarity 47.8%; Pred. No. 0.044;
Matches 164; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 26 CAAATACATGATGACAAAGTCCCAAAAAATTTTGGTACAGAAACCTCCAAATTTT 85
DB 4771 CAAAAATAAAATAAAATAATAATAATAAAATCTAAATAATCATATAAAATATCTT 4712
QY 86 TTTTTTATGGACAAAGAAATACAGATAGAAACATTTTGTGTGGATGGAAGTACT 145
DB 4711 TTAATTAATAAACTAAAAAACCCTAAATTTATATATACAAAAATCAAAAAATTTT 4652
QY 146 AATATACATTAAGCAATTTTAAAAAATATATATAGCCCTATAGCCCTCAAAAGTATGTTA 205
DB 4651 AACTTAACTATATAAACTCATATACTAAAAACGTAATAAAATAAACTAAATAAAA 4592
QY 206 TCTAGTAGTGTAATTAATGATGCGGATTCAGAAATGGGACACAAATGAACG 265
DB 4591 TAAATAAAAAATAAATAATTCATATATATATATATATATATATATATATATATAT 4532
QY 266 GAATTAATAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 325
DB 4531 TAAACTACTAATTTAAAAAATAACGAAACTACTTATATATATATATATATATATAT 4472

QY 326 GAGGGCAAAAAAGACAAATGCCAAAGCTCTACGGTTTGAC 368
DB 4471 TTTATCAAAATAAATCATTTCAAAATAATATACATTTTATAAC 4429
RESULT 3
AAS61342/C
ID AAS61342 standard; DNA; 6107 BP.
XX AAS61342;
XX 29-JAN-2002 (first entry)
XX Human gene regulation-associated gene oligonucleotide #297.
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX WO200177375-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-EP03968.
XX 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017470/02.
XX New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease
XX Disclosure; SEQ ID NO 304; 26pp; English.
XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences) The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases; by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;

Query Match	5.1%	Score 56.6	DB 24	Length 6107
Best Local Similarity	47.8%	Pred. No. 0.044		
Matches 164	Conservative 0	Mismatches 179	Indels 0	Gaps 0
QY	26	CAATAACAAATGCATAGAAAGTTCCAAAAAATATTTGTTAAACAGAAAACCTTCCAAATTTT	85	
DB	4771	CAAAAAAATAAATAAATAATATATATATAAAAAAATCTTAAAAATACATAAAATATACCTT	4712	
QY	86	TTTTTTTTATGGAAACAAGAAATACAGATAGAAAACATATTTCTTGTGGAATCGAAAGTAGT	145	
DB	4711	TTAAATAAAAAAATAAAAAAACCCACATAAAATTTATAATTTACAAAAATCAAAAATTTT	4652	
QY	146	AATATACATTAAGCAAAATTTTAAAAAATATATATAAGCCTATACGGCTCAAAAGTAGTTA	205	
DB	4651	AAACTTAACTATATAAAACTCATATAACTTAAAAAAAACGTAAAAAATAAACTAAAAATAAA	4592	
QY	206	TCTAGTAGGTGTAATTAATTAATGCAATGGTCCGATTCAGAAATTTGGGACACAAATGAAACG	265	
DB	4591	TAAAAATAAAAAAATAAATTAATTCATATATCATATAACAATAAACTTAACTATCATAAAT	4532	
QY	266	GAAHTTAAATATTAACCTTTAAATTAATTAATAATTTTGAGTAAATGTGTTCTCGACTATT	325	
DB	4531	TAAAAACTCATATTTTAAAAAATAAACGAAACACTATTATATTATCATCACTTTCTTTCC	4472	
QY	326	GAGGGGCAAAAAAGACAAATGCCAAAAAGCTACGGGTTTGAC	368	
DB	4471	TTTAATCAAAAATAAATCATTTTCAAAAAATAATACATTTTATAC	4429	

RESULT	4
ABK31431/c	
ID	ABK31431 standard; DNA; 6107 BP.
XX	
XX	
AC	ABK31431;
XX	
XX	
DT	23-APR-2002 (first entry)
XX	
XX	Signal transduction associated gene modified complementary DNA #137.
DE	
XX	
XX	Human; signal transduction associated gene; cytosine methylation state;
KW	CpG island; signal transduction associated disease; solid tumour; cancer;
KW	antitumour; cytostatic; mutant; ds.
KW	

CC	epigenetic parameters of genes associated with signal transduction.
CC	The genomic DNA can be obtained from cells or cellular components which
CC	contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC	cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC	eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC	histologic object slides, and all their possible combinations. The
CC	sequences of the invention are useful for the diagnosis and therapy of
CC	diseases associated with signal transduction e.g. solid tumours and
CC	cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC	sequences of different genes associated with signal transduction, or
CC	their complementary sequences.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SQ	Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;
	Query Match 5.1%; Score 56.6; DB 24; Length 6107;
	Best Local Similarity 47.8%; Pred. No. 0.044;
	Matches 164; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY	26 CAATTAACAATGCATGAAGAAGTCCAAAAAATTTTGTGTAAACAGAAACCTTCCAAATTTT 85
DB	4771 CAAAAAATAAATAAATAAATAATATATATAAAAAATCTAAAAAAATACATNAATATCCCT 4712
QY	86 TTTTTTTTATGGAACAAGAAATTAACAGATAGAAAACTATTTTGTGTGCAATGGAAGTAGT 145
DB	4711 TTAATAAAAAAATAAATAAATAAATAAATAAATTTATTAATACAAAAATCAAAAAATTT 4652
QY	146 AATATACATTAAGCAAAATTTTAAAAAATATATATAAGCCTATACGGCTCAAGTATGTTA 205
DB	4651 AAACCTTAACATATATAAACTCATATAAATAAATAAATAAATAAATAAATAAATAAATAA 4592
QY	206 TCTAGTAGGTGTAATTAATATGATGCGATTCAGAAATTCGACAAATGGGACAACAATGAAACG 265
DB	4591 TAAAAATAAAAAAATAAATTAATTCATATATATCATACAAAAAATTTACTTAATCTCTAAAT 4532
QY	266 GAATTTAAATATTTAACTTTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAA 325
DB	4531 TAAAACTCATAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4472
QY	326 GAGGGGCAAAAAAGACAAATGCCAAAAAGCTACGGGTTTGAC 368
DB	4471 TTTTAATCAAAATAAATAATCATTTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 4429

RESULT 5	
AAA01920	
ID	AAA01920 standard; cDNA; 875 BP.
XX	
XX	AAA01920;
XX	
XX	19-MAY-2000 (first entry)
XX	
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:1911.
XX	
XX	Human; colon cancer; tumour; diagnosis; gene expression product;
KW	probe; detection; cancerous state; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.

```

PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
XX cells -
XX
XX Claim 1; Page 720; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
XX libraries constructed from human colon cancer cell lines. The present
XX invention also describes a method of detecting differentially expressed
XX genes correlated with a cancerous state of a mammalian cell, comprising
XX detecting at least one differentially expressed gene product in a test
XX sample derived from a cell suspected of being cancerous, where detection
XX of the differentially expressed gene product is correlated with a
XX cancerous state of the cell from which the test sample was derived.
XX The polynucleotide sequences can be used in a method for detecting
XX differentially expressed genes correlated with a cancerous state of a
XX mammalian cell. The polynucleotides can also be used as probes for
XX detecting and mapping related genes. They can be used in diagnosis and
XX prognosis of diseases and disorders (e.g. identification of
XX pre-metastatic or metastatic cancerous states, stages of cancer, or
XX responsiveness of cancer to therapy). This is particularly for breast
XX cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
XX negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 875 BP; 421 A; 10 C; 78 G; 198 T; 168 other;
XX
XX Query Match 4.9%; Score 54.2; DB 21; Length 875;
XX Best Local Similarity 42.6%; Pred. No. 0.1;
XX Matches 149; Conservative 0; Mismatches 196; Indels 5; Gaps 1;
XX
XX 1 AATCTACAAATTTATTAATATTTACTCAATCAACATGATGAGAAAGTTCACAAAATTT 60
XX III III III III III III III III III III III III III III
XX 525 AATANTNATATATANTANNAAAAATATATAAAATAAGAGATATGNNAAANGAAAAA 584
XX
XX 61 TGTTCACAGAACTTCCAAATTTTTTTTTTTATGGAACAAGAAATAACAGATAGAAAC 120
XX II III III III III III III III III III III III III III
XX 585 ATATANAGGAAAGTAAATTAATNGATATTTAAGANAAAGAAAAAATAATANAN 644
XX
XX 121 TATTTTGTCTGCAATGGAAGTAGTAAATATACATTAAGCAAAATTTTAAAAAATTTATATA 180
XX II III III III III III III III III III III III III III
XX 645 NATNANAATATANTAAANTNAAANTANANAANAATNCNAAATNTANNAGATNATAAGA 704
XX
XX 181 GCCTATACGGCTCA-----AAGTAGTTTATCTAGTAGGTGTAATTAATTAATGCGATGGTG 235
XX II III III III III III III III III III III III III III
XX 705 NAANNTATNAAATNAACTNAGAAATATATNAAANNATAGNACTTANANTNNTATAAGAN 764
XX
XX 236 CGATTTCAGAAATGGGACACATGAAACCGGAATTAATAATTAATTAATTAATTAATTAATTA 295
XX II III III III III III III III III III III III III III
XX 765 GNNTTAAGGANAATANATNAAATAGATANNNAATNAAATTAATTAATTAATTAATTAATTA 824
XX
XX 296 AAATTTAGTAATAATGTGTTTCTGACTATTGAGGGGCAAAAAAGACAA 345
XX II III III III III III III III III III III III III III
XX 825 AATAATGAGTGANNGATTATNAAANNATNANNCTATNTAATATAATAA 874
XX
XX
XX RESULT 6
XX ABL32803/c
XX ID ABL32803 standard; DNA; 6113 BP.
XX AC ABL32803;
XX XX
XX DT 26-MAR-2002 (first entry)
XX

```

```

XX Human immune system associated gene SEQ ID NO: 776.
XX DE
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
XX antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
XX OS
XX XX WO200200928-A2.
XX PN
XX XX 03-JAN-2002.
XX PD
XX XX 02-JUL-2001; 2001WO-EP07537.
XX PF
XX XX 30-JUN-2000; 2000DE-1032529.
XX PR
XX XX 01-SEP-2000; 2000DE-1043826.
XX PR
XX XX (EPIG-) EPIGENOMICS AG.
XX PA
XX XX Olek A, Piepenbrock C, Berlin K;
XX PI
XX XX WPI; 2002-130909/17.
XX DR
XX XX
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation
XX PT
XX XX Claim 1; SEQ ID NO 776; 32pp + Sequence Listing; German.
XX PS
XX XX
XX CC The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX CC
XX
XX Sequence 6113 BP; 1694 A; 157 C; 1445 G; 2817 T; 0 other;
XX
XX Query Match 4.9%; Score 54.2; DB 24; Length 6113;
XX Best Local Similarity 51.9%; Pred. No. 0.13;
XX Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
XX
XX 6 ACAAAATTATTAATTTAGTCAATAACAATGCATAGAAGTTCACAAAATTTTGTGA 65
XX III III III III III III III III III III III III III III
XX 5823 ACACAAACTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5764
XX
XX 66 ACAGAAACTTCCAAATTTTTTTTTTTTATGGAACAAGAAATAACAGATAGAAATTTT 125
XX III III III III III III III III III III III III III III
XX 5763 AAAAAAATTTTAAATATATATATATTAATAAATAAATAAATAAATAAATAAATAA 5704
XX
XX 126 TGTGTGGAATGGAAGTAGTAATATACATTAAGCAAAATTTTAAAAAATTAATAAGCCTA 185
XX III III III III III III III III III III III III III III
XX 5703 TCCTTTTAAAAAATTCATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5644
XX
XX 186 TAGCGGCTCAAGTATGTTTATCTAGTAGGTGTAATTAATTAATTAATTAATTAATTAAT 240
XX III III III III III III III III III III III III III III
XX 5643 CGTATCTATAAATCTTTTTTTTTTTTTTTTTTTTTTAAATAAATCTCGCCCTATT 5589
XX
XX
XX RESULT 7
XX ABL34197/c
XX ID ABL34197 standard; DNA; 37973 BP.
XX AC ABL34197;
XX XX
XX DT

```

DT	26-MAR-2002	(first entry)
XX	Human immune system associated gene SEQ ID NO: 2170.	
XX	Human; Immune system disease; cytosine methylation; antiasthmatic;	
KW	antiarteriosclerotic; antianemic; cytostatic; nootropic;	
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;	
KW	kneuropathic; antiarthritic; antidiabetic; antipsoriatic;	
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
KW	gene; ds.	
XX	Homo sapiens.	
XX	WO200200928-A2.	
PN	03-JAN-2002.	
XX	02-JUL-2001; 2001WO-EP07537.	
XX	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
XX	(EPIG-) EPIGENOMICS AG.	
PA	Olek A, Piepenbrock C, Berlin K;	
XX	WI; 2002-130909/17.	
DR	Nucleic acid comprising fragment of chemically modified gene, useful	
PT	for diagnosis and treatment of diseases associated with abnormal	
PT	cytosine methylation	
XX	Claim 1; SEQ ID NO 2170; 32pp + Sequence Listing; German..	
PS	The present invention provides a number of human immune system associated	
CC	genes which are modified by the methylation of cytosines. The sequences	
CC	can be used in the diagnosis and treatment of immune system disorders,	
CC	including eye diseases such as retinopathy, neovascular glaucoma and	
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid	
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,	
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel	
CC	diseases. The present sequence is a genome of the invention.	
XX	Sequence 37973 BP; 10756 A; 620 C; 8310 G; 18287 T; 0 other;	
SQ	Query Match 4.9%; Score 53.8; DB 24; Length 37973;	
	Best Local Similarity 50.2%; Pred. No. 0.2;	
	Matches 159; Conservative 0; Mismatches 157; Indels 1; Gaps 1;	
QY	8 AAATTATTAATATTTAGTCAATAACAATGCTAGAGAATGCCAAAAAAATTTTGTTAAC 67 	
Db	8417 AAAAATAAATCTTAAAAAAATCTAATACAAAATAATAAAACACATNAATTTACTAAC 8358	
QY	68 AGAAATCTCCAAATTTTTTTTTTTATGGGAACAAGA-AAATAACAGATAGAAAACATATTT 126 	
Db	8357 ATAAAAAACAATAAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8298	
QY	127 GTTGTGGAATGGAAGTAGTAGTAATATACATTAAGCAAATTTTAAAAAATATATATAAGCCTAT 186 	
Db	8297 AAATTAATAAAAAATTAATAAACAATAATTAATAAATTTAACATAATAACAAAAAATATT 8238	
QY	187 ACGCGCTCAAGTAGTGTATCTAGTAGGTGAATTAATATATGCATGGTCGANTCAGAAT 246 	
Db	8237 AAAAACTAAAAAATAAATAAATTTACTACTAATAATTAATTAATAAATAAATAAATAA 8178	
QY	247 TGGGACAACAATGAAACGGAATTAATAATTAACCTTAAAAATAAATAAATAAATTTGAGTA 306 	
Db	8177 AAAAAATACAAATTAATAATTTTACAAAAAACAATTTCAAAATTTATACATCTTTTTT 8118 	
QY	307 AATGCTGTTTCTGACTA 323 	


```
Db 27154 ACATACCTTATAGTATATATTTAAATATATATGTAATGATATATTTTAAATGATGATATA 27213
Qy 245 ATTGGGACACAATGAAACCGAATTTAAATATTTAACTTTTAAATATAAATAAATTTTGGAG 304
Db 27214 TAATATACATTTTAAATACACATTTATATTTATTTATATAAATATATATAAATCTCC 27273
Qy 305 TAAATGCTTTTC 317
Db 27274 AAGTTGCTTTTC 27286

RESULT 14
ABA82623
ID ABA82623 standard; DNA; 72049 BP.
XX
XX ABA82623;
XX
XX 25-JAN-2002 (first entry)
XX
XX Human HBM gene region b527d12-h_contig309G.
XX
XX Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
XX sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
XX antisense therapy; vaccine; bone disorder; Paget's disease;
XX sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
XX Homo sapiens.
XX OS
XX WO200177327-A1.
XX PN
XX 18-OCT-2001.
XX PD
XX
XX 21-JUN-2000; 2000WO-US16951.
XX PF
XX 05-APR-2000; 2000US-0543771.
XX PR
XX 05-APR-2000; 2000US-0544398.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX PA
XX
XX Carulli JP, Little RD, Recker RR, Johnson ML;
XX PI
XX WPI; 2001-657171/75.
XX DR
XX
XX New high bone mass (HBM) and Zmax1 genes and proteins useful for
XX modulating bone mass for the treatment of e.g. osteoporosis -
XX
XX Claim 51; Page 258-302; 443pp; English.
XX
XX The present invention describes the human Zmax1 gene and the high bone
XX mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
XX HBM genes have osteopathic activities. The genes can be used in gene
XX therapy, antisense therapy and in the production of vaccines. They
XX can be used in the diagnosis and treatment of bone disorders including
XX osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
XX dysplasia. ABA82038 to ABA82700 and ABA88188 to ABA88193 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 72049 BP; 15164 A; 19306 C; 20142 G; 17434 T; 3 other;
XX SQ

Query Match 4.88; Score 53; DB 22; Length 72049;
Best Local Similarity 50.2%; Pred. No. 0.32; Indels 1; Gaps 1;
Matches 157; Conservative 0; Mismatches 155;

Qy 6 ACAAAATTAAATTTAGTCAATAACATGCAATGCAAGTTCACAAAATTTTCTTA 65
Db 30562 AATATATAAATATATATTTTAAATATATATATATATATATAAATATATATATATA 30621
Qy 66 ACAGAAACTCCAAATTTTTTTTTTTTATGGA-ACAAGAAATACAGATAGAAACTATT 124
Db 30622 AATAAAATTTTATAAATATATATTTTATAAGTAAATATAATATAAATATAAATATA 30681
Qy 125 TTCTGTGGAGTGAAGTACTATATACATTAACCAATTTTAAAAAATTTATATAACCT 184
Db 125 TTCTGTGGAGTGAAGTACTATATACATTAACCAATTTTAAAAAATTTATATAACCT 184

Db 30682 TTATATAATATATAATAAATAAATATATAAATAAATAAATATATTTTAAATAATATATAAT 30741
Qy 185 ATACCGCTCAAGTAGTATCTATCTAGTCTAATTAATGATGATGCGCATTCAGA 244
Db 30742 ACATACCTTATAGTATATATTTAAATATATATGTAATGATATATTTTAAATGATGATATA 30801
Qy 245 ATTGGGACACAATGAAACCGAATTTAAATATTTAACTTTTAAATATAAATAAATTTTGGAG 304
Db 30802 TAATATACATTTTAAATACACATTTATATTTATTTTATATAAATATATAAATAATCTCC 30861
Qy 305 TAAATGCTTTTC 317
Db 30862 AAGTTGCTTTTC 30874

RESULT 15
ABL70459/C
ID ABL70459 standard; DNA; 16228 BP.
XX
XX ABL70459;
XX
XX 01-JUL-2002 (first entry)
XX
XX Chemically treated cell signalling DNA sequence#175.
XX
XX Cell signalling; cytosine methylation; cell signalling disease;
XX cancer; tumour; cytostatic; ds.
XX
XX Unidentified.
XX OS
XX WO200202807-A2.
XX PN
XX 10-JAN-2002.
XX PD
XX
XX 29-JUN-2001; 2001WO-EP07471.
XX PF
XX 30-JUN-2000; 2000DE-1032529.
XX PR
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX PA
XX
XX Olek A, Piepenbrock C, Berlin K;
XX PI
XX WPI; 2002-154758/20.
XX DR
XX
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signalling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signalling -
XX
XX Claim 1; SEQ ID NO 349; 24pp+sequence listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
XX 18 bases of a segment of chemically pretreated DNA of genes associated
XX with cell signalling. The activity of the modified sequences of the
XX invention may be described as cytostatic. The object of the invention is
XX to provide the chemically modified DNA of genes associated with cell
XX signalling, as well as oligonucleotides and/or PNA-oligomers for
XX detecting cytosine methylations, as well as a method which is
XX particularly suitable for the diagnosis and/or therapy of genetic and
XX epigenetic parameters of genes associated with cell signalling. The
XX chemically modified DNA provided by the invention is useful for diagnosis
XX and therapy of diseases such as solid tumours and cancer. The sequences
XX given in records ABL70111-ABL70626 represent chemically pre-treated
XX genomic DNA's of genes associated with cell signalling.
XX
XX Note: The sequence data for this patent is not represented in the printed
XX specification, but is based on sequence information supplied by the
XX European Patent Office.
XX
XX Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;
XX SQ

Query Match 4.88; Score 52.6; DB 24; Length 16228;
Best Local Similarity 48.2%; Pred. No. 0.31; Indels 0; Gaps 0;
Matches 148; Conservative 0; Mismatches 159;
```

QY		47	TCGAAATAAATTTTGGTTACAGAACTTCCAAAATTTTTTTTTTTATGGAACAAGAAT	106
Db		7878	TCTACTATTAAACTATTATTTAAAAAACAAACTTTTTTCTTTTTTAAACGAAATA	7819
QY		107	AACAGATAGAAAACACTATTTTGTTGGAAATGGAAGTAGTAATATACATTAAAGCAATTTT	166
Db		7818	TACCGTAACATAATTATTTCCCAANTTCTACTTAATACCTCATAACTTAATAAAAAACA	7759
QY		167	AAAAAATTTATAAGCCTATACGCCGTCAAAGTGTCTATCTAGTAGGTTGAATATAA	226
Db		7758	AAAAAATAAATAAATAAATAAAGAAATACACTTATAAATATTTACTTCCCTAT	7699
QY		227	TGCATGGTGGATTCAGAATTTGGGACAACAATCAAAACGGAATTAATAATTAACTTAA	286
Db		7698	TAAACATAAATAAATAAATAACCATAAANAATCAANATATAAATTAATAAATTA	7639
QY	.	287	AATAAATAAATAATTTGAGTAATGTCTTTCTGACTATTGAGGGGCAAAAAAAGACAAT	346
Db		7638	AACTATTCTTAATAATAACAATATCTTATAAAAATAAATACAATTAATAAATAATTT	7579
QY		347	GCCAAA	353
Db		7578	TAATAAA	7572

Search completed: June 10, 2003, 22:22:43
Job time : 254.042 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 52.4398 Seconds
(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202
Perfect score: 1101
Sequence: 1 aatcacaaattattaat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2.6/prodata/1/lna/5A-COMB.seq:*
2: /cgn2.6/prodata/1/lna/5B-COMB.seq:*
3: /cgn2.6/prodata/1/lna/6A-COMB.seq:*
4: /cgn2.6/prodata/1/lna/6B-COMB.seq:*
5: /cgn2.6/prodata/1/lna/PCTUS-COMB.seq:*
6: /cgn2.6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	49.2	4.5	3350	3	US-08-617-860B-3
C 2	49.2	4.5	6265	4	US-09-129-112-3
C 3	46.6	4.2	19124	2	US-08-487-826B-13
C 4	46.4	4.2	636	4	US-08-998-416-1137
C 5	46	4.2	731	1	US-08-451-405A-2
C 6	45.8	4.2	665	2	US-08-883-795A-36
C 7	45.8	4.2	51952	3	US-08-947-823-1
C 8	45	4.1	20674	4	US-08-641-638-651
C 9	44.8	4.1	3138	1	US-07-867-106-4
C 10	44.6	4.1	19124	2	US-08-487-826B-13
C 11	44	4.0	7218	1	US-08-232-463-14
C 12	43.8	4.0	6124	4	US-08-213-419B-3
C 13	43.6	4.0	5852	1	US-07-867-106-2
C 14	43.2	3.9	660	1	US-07-991-867B-32
C 15	43.2	3.9	660	1	US-08-107-755A-32
C 16	43.2	3.9	660	2	US-08-544-332-32
C 17	43.2	3.9	660	4	US-09-370-861A-32
C 18	43.2	3.9	660	4	US-07-991-867B-8
C 19	43.2	3.9	1511	1	US-08-107-755A-8
C 20	43.2	3.9	1511	1	US-08-544-332-8
C 21	43.2	3.9	1511	4	US-09-370-861A-8
C 22	43.2	3.9	4810	3	US-08-852-629-11
C 23	43.2	3.9	4838	3	US-08-852-629-15
C 24	42.8	3.9	1850	3	US-08-617-860B-32
C 25	42.8	3.9	4098	2	US-08-605-106-4
C 26	42.4	3.9	470	4	US-09-020-956-102
C 27	42.4	3.9	470	4	US-09-030-607-102

C 28	42.4	3.9	470	4	US-09-605-785-102	Sequence 102, App
C 29	42.4	3.9	470	4	US-09-439-313-102	Sequence 102, App
C 30	42.4	3.9	470	4	US-09-352-616A-102	Sequence 102, App
C 31	42.4	3.9	470	4	US-09-232-149A-102	Sequence 102, App
C 32	42.4	3.9	615	4	US-08-998-416-186	Sequence 186, App
C 33	42.4	3.9	5181	1	US-08-257-073-10	Sequence 10, Appl
C 34	42.4	3.8	1406	3	US-08-913-842-6	Sequence 6, Appl
C 35	41.8	3.8	658	4	US-08-998-416-595	Sequence 595, App
C 36	41.4	3.8	2861	1	US-08-299-953-1	Sequence 1, Appl
C 37	41.4	3.8	2861	1	US-08-459-415-1	Sequence 1, Appl
C 38	41.4	3.8	2861	4	US-09-066-687-1	Sequence 1, Appl
C 39	41.4	3.8	2861	5	PCT-US95-11231-1	Sequence 1, Appl
C 40	41.4	3.8	3881	1	US-08-299-953-2	Sequence 2, Appl
C 41	41.4	3.8	3881	1	US-08-459-415-2	Sequence 2, Appl
C 42	41.4	3.8	3881	4	US-09-066-687-2	Sequence 2, Appl
C 43	41.4	3.8	3881	5	PCT-US95-11231-2	Sequence 63, Appl
C 44	41	3.7	1441	4	US-08-821-994-63	Sequence 1, Appl
C 45	41	3.7	2110	4	US-09-419-459-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-617-860B-3/c
: Sequence 3, Application US/08617860B
: Patent No. 6133506
: GENERAL INFORMATION:
: APPLICANT: Tvpfer, R., Bautor, J., Bothmann, H., Filsak, E.,
: APPLICANT: Hrvicke-Grandpierre, C., Klein, B., Martini, N.,
: APPLICANT: M ller, A., Schulte, W., Voetz, M., Walek, J.,
: APPLICANT: Schell, J.
: TITLE OF INVENTION: Promoters
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS: 35
: ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
: STREET: 1140 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: New York
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,860B
: FILING DATE: 01-MAR-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP94/02950
: FILING DATE: 05-SEP-1994
: APPLICATION NUMBER: DE F4329951.2
: FILING DATE: 04-SEP-1993
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3350 Base pairs
: TYPE: Nucleic acid
: STRANDEDNESS: Double stranded
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Brassica napus
: IMMEDIATE SOURCE:
: LIBRARY: genomic Lambda FIX II
: CLONE: BnACCasg10
: FEATURE:
: NAME/KEY: Startcodon
: LOCATION: 2611..2613
: FEATURE:
: NAME/KEY: CDS

;
US-08-617-860B-3

Query Match	4.5%	Score 49.2;	DB 3;	Length 3350;
Best Local Similarity	48.7%;	Pred. No. 0.02;		
Matches 190;	Conservative 0;	Mismatches 198;	Indels 2;	Gaps 2;
QY	6	ACAAATATTAAATATTTAGTCAATAACAATGCATAGAAAGTCCCAAAAAAATTTTCCTTA	65	
Db	2312	ATAAATGAATTAATTTTATATATATATATATAT -CAGCTATGTGAATAAAAAATATAGAA	2254	
QY	66	ACAGAAATCTCCAAATTTTTTTTTTTATGGAAACAAGAAATAACAGATAGAAAACATATTT	125	
Db	2253	ACTAAAAATTAAATAATTTAAATGTTTTTAAACATAAAATAATAGTTTGGGTA -TGTAT	2195	
QY	126	TGTTGTGGGAATGGAAGTAGTAATAATACATTAACCAATTTTAAAAAATATATATAAGCCTA	185	
Db	2194	ATATTTTATGTTTAAAAACATTTAGAAAAATATTTAAATTTAGTTTAAAAATTTTATTTTC	2135	
QY	186	TACGCGCTCAAAATATGTTATCTAGTAGGTGTAAATTAATAATGCATGGTCCGATTACAGAA	245	
Db	2134	ATTTGACATACAAAATATCAAAAATAGTTTAGACTATTTATAAATTTTCTAACAAAGT	2075	
QY	246	TTGGGACACAATGAAAACGGAATTTAAATATTAACTTTAAAAATAAATAAAAAATTTGAGT	305	
Db	2074	AAGAGTTATGATATATAAAAAATCAAGACATAAAAATTTATAAATATTTAAAAATTTTAAT	2015	
QY	306	AAATGTGTTTCTTGACTATTGTAGGGGCAAAAAAAGACAAATGCCAAAGTCTACGGGTTT	365	
Db	2014	GTCAATATATAATTTAACTTTTAATACAAATTAACCAAAAAATAAAGATCATTTGAATA	1955	
QY	366	GACTGTCCAGTTCGGTAATATCTAATAAC	395	
Db	1954	AATGTGCAATAACAGAAATAAATCAACACC	1925	

```

RESULT 2
US-09-129-112-3/c
; Sequence 3, Application US/09129112
; Patent No. 6465716
; GENERAL INFORMATION:
; APPLICANT: Etzler, Marilynn E.
; APPLICANT: Murphy, Judith B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
; FILE REFERENCE: 023070-079810US
; CURRENT APPLICATION NUMBER: US/09/129,112
; CURRENT FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 08/907,226
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 6265
; TYPE: DNA
; ORGANISM: Dolichos biflorus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (633)..(944)
; NAME/KEY: intron
; LOCATION: (945)..(1022)
; NAME/KEY: exon
; LOCATION: (1023)..(1151)
; NAME/KEY: intron
; LOCATION: (1152)..(1559)
; NAME/KEY: exon
; LOCATION: (1560)..(1616)
; NAME/KEY: intron
; LOCATION: (1617)..(1697)
; NAME/KEY: exon
; LOCATION: (1698)..(1790)
;
; OTHER INFORMATION: genomic sequence of NBP46 (DB46)
US-09-129-112-3

```

RESULT 3
 US-08-487-826B-13
 Sequence 13, Application US/08487826B
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellemis, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelsen, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match
Best Local Similarity 4.2%; Score 46.6; DB 2; Length 19124;
Matches 139; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 9 AATTATTAAATTTAGTCAATCAATGCGATAGAGTTCCCAAAAAATTTTGTAAACA 68
Db 15735 AAAAAATATATCATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15794
QY 69 GAAACTCCCAATTTTATTTTATGGAAGAATAACAGATAGAGAACTATTTTCT 128
Db 15795 TAAATAAATAAATTTTATTTTAAATAAATAAATAAATAAATAAATAAATAA 15854
QY 129 TGTGGAATGGAAGTAGTAAATATACATTAAGCAATTTTAAATAAATAAATAAATAA 15854
Db 15855 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15914
QY 189 GCGCTCAAGTATGTTATCTAGTGTAGTGTAAATTAATGCGATGCGATTCAGAAATG 248
Db 15915 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 248
QY 249 GGACAAATGAAACGGAATTAATAATATTAATTAATTAATTAATTAATTAATTAAT 15974
Db 15975 GCACATACATACAT 16027

RESULT 4

US-08-998-416-1137/c
Sequence 1137, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:
APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:

Query Match
Best Local Similarity 4.2%; Score 46.6; DB 2; Length 19124;
Matches 139; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 9 AATTATTAAATTTAGTCAATCAATGCGATAGAGTTCCCAAAAAATTTTGTAAACA 68
Db 15735 AAAAAATATATCATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15794
QY 69 GAAACTCCCAATTTTATTTTATGGAAGAATAACAGATAGAGAACTATTTTCT 128
Db 15795 TAAATAAATAAATTTTATTTTAAATAAATAAATAAATAAATAAATAAATAA 15854
QY 129 TGTGGAATGGAAGTAGTAAATATACATTAAGCAATTTTAAATAAATAAATAAATAA 15854
Db 15855 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15914
QY 189 GCGCTCAAGTATGTTATCTAGTGTAGTGTAAATTAATGCGATGCGATTCAGAAATG 248
Db 15915 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 248
QY 249 GGACAAATGAAACGGAATTAATAATATTAATTAATTAATTAATTAATTAATTAAT 15974
Db 15975 GCACATACATACAT 16027

LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match
Best Local Similarity 4.2%; Score 46.4; DB 4; Length 636;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 46 TTCACAAAATAATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 105
Db 636 TTTTATAGTATTTTAACTACATCTTTATATATATTTTATATATATTTTATATATATAT 577
QY 106 TAACAGATAGAAAATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 165
Db 576 AAAATATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 517
QY 166 TAAAAATTAATTAAGCCTATACGCTCAAGTATGTTATCTAGTGTAGTGTAGTGTAGTGT 225
Db 516 TATTAATATATATAAATAATTAATAAAGAAATTAAGTAAATTAATTAATTAATTAATTAAT 457
QY 226 ATGCATGGTCCGATTCAGAAATGGGACAAATGGAATGGAATGGAATGGAATGGAATGGAAT 285
Db 456 TTCTTAAAAAGATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 397
QY 286 AATAAATAAATAATTT 301
Db 396 ATAAAAATAATAATTT 381

RESULT 5

US-08-451-405A-2
Sequence 2, Application US/08451405A
Patent No. 5736358

GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE WEBB LAW FIRM
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: Midwest Micro 486-50
OPERATING SYSTEM: DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,405A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,273
FILING DATE: 15-JAN-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match
Best Local Similarity 4.2%; Score 46; DB 1; Length 731;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 46 TTCACAAAATAATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 105
Db 636 TTTTATAGTATTTTAACTACATCTTTATATATATTTTATATATATTTTATATATATAT 577
QY 106 TAACAGATAGAAAATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 165
Db 576 AAAATATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 517
QY 166 TAAAAATTAATTAAGCCTATACGCTCAAGTATGTTATCTAGTGTAGTGTAGTGTAGTGT 225
Db 516 TATTAATATATATAAATAATTAATAAAGAAATTAAGTAAATTAATTAATTAATTAATTAAT 457
QY 226 ATGCATGGTCCGATTCAGAAATGGGACAAATGGAATGGAATGGAATGGAATGGAATGGAAT 285
Db 456 TTCTTAAAAAGATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 397
QY 286 AATAAATAAATAATTT 301
Db 396 ATAAAAATAATAATTT 381


```
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 51952 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

Query Match
Best Local Similarity 4.2%; Score 45.8; DB 3; Length 51952;
Matches 170; Conservative 0; Mismatches 182; Indels 5; Gaps 1;

QY 1 AATCTCAAAATTATTAATTTAGTCAATAAACAATGCATAGAAAGTTCACAAAATAATTT 60
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 TGTAAACAGAACTTCCAAATTTTTTTTATGGAACAAGAAATAACAGATAGAAAAC 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 TATTTTGTGGAATGGAAGTAGTAATATACATTAAGCAAAATTTTAAAAAATTATATAA 180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GCCTATACG-----CGCTCAAGTAGTGTATCTAGTGTGTAATTAATAATGATGTG 235
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 236 CGATTCAAGATTTGGGACAAATGAAACGGAAATTAATTAATTAATTAATTAATTAATTA 295
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 296 AAATTGAGTAATGTGTTCTGACTATTGAGGGGCAAAAAGACAAATGCCAAA 352
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 32833 TTTCTGTATATATAATATGATAAACTATTTAACAGGGGAAATTTGACAGAAAAA 32889
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-09-641-638-651/c
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1123..3123
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon

; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6
; NAME/KEY: exon
; LOCATION: 7379..7522
; OTHER INFORMATION: exon 7
; NAME/KEY: exon
; LOCATION: 8645..8854
; OTHER INFORMATION: exon 8
; NAME/KEY: exon
; LOCATION: 12254..12340
; OTHER INFORMATION: exon 9
; NAME/KEY: exon
; LOCATION: 12854..13023
; OTHER INFORMATION: exon 10
; NAME/KEY: exon
; LOCATION: 13308..13429
; OTHER INFORMATION: exon 11
; NAME/KEY: exon
; LOCATION: 16567..16667
; OTHER INFORMATION: exon 12
; NAME/KEY: exon
; LOCATION: 16775..16945
; OTHER INFORMATION: exon 13
; NAME/KEY: exon
; LOCATION: 17063..17554
; OTHER INFORMATION: exon 14
; NAME/KEY: misc_feature
; LOCATION: 17555..20674
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 1128
; OTHER INFORMATION: 10-508-191 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1182
; OTHER INFORMATION: 10-508-245 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
```

```
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
```

```
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 4.1%; Score 45; DB 4; Length 20674;
Best Local Similarity 45.9%; Pred. No. 0.27;
Matches 153; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 1 AATCTACAAATTAATTAATTTAGTCATAAATCCATAGAAAGTCCCAAAAAAATTT 60
Db 11391 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTA 11392

QY 61 TGTTAACAGAACTTCCAAATTTTTTTTTTATGGAACAAGAAATAACAGATAGAAAAC 120
Db 11331 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11272

QY 121 TATTTTGTGTGGAATGGAAGTAGTAATATATACATTAAAGCAAAATTTAAAAAATATATAA 180
Db 11271 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11212

QY 181 GCCTATACGGCTCAAGTAGTCTATCTAGTAGTGTATTAATTAATGAATGCGGATT 240
Db 11211 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAAAT 11152

QY 241 CAGAAATGGCAACAATGAAACCGAATTAATAATTAATACTTTAAAAATAAAAAATTT 300
Db 11151 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11092

QY 301 TGAGTAATGTGTTTCTGACTATTGAGGGCA 333
Db 11091 TAAATATTTTCCTTAGCTATTAAAAAAGGCA 11059
```

RESULT 9

```
US-07-867-106-4/c
; Sequence 4, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved plasmid vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


Search completed: June 11, 2003, 01:53:47
Job time : 54.4398 secs

TYPE: DNA
ORGANISM: Dolichos biflorus
FEATURE:
OTHER INFORMATION: genomic sequence of NBP46 (DB46)
NAME/KEY: exon
LOCATION: (633)..(944)
NAME/KEY: Intron
LOCATION: (945)..(1022)
NAME/KEY: exon
LOCATION: (1023)..(1151)
NAME/KEY: Intron
LOCATION: (1152)..(1559)
NAME/KEY: exon
LOCATION: (1560)..(1616)
NAME/KEY: Intron
LOCATION: (1617)..(1697)
NAME/KEY: exon
LOCATION: (1698)..(1790)
US-09-129-112-3

Query Match
Best Local Similarity 4.5%; Score 49.2; DB 10; Length 6265;
Matches 173; Conservative 0; Mismatches 148; Indels 5; Gaps 3;
QY 15 TAATATTAGTCATACATGCTAGAGAAAGTTCCAAAATAATTTTGTAAACAGAAACT 74
DB 2936 TATTTATAGTGTCTTACTATTCAATAAACAATAATGCGAGATATGAAGTTTACTG-AACT 2878
QY 75 TCCAAATTTTTTTTTTATGGAACAAGAAATAACAGATAGAAACTATTTTGTGTGA 134
DB 2877 TCTTTTTTTTTTTTTTATTGAAATGTTTGAATATGTCGAATATTAATATAGTGA 2818
QY 135 ATCGAGTAGTATATACATTTAA--GCAATTTTAAAAATTTATATAAGCCTATACGCGC 192
DB 2817 ATATTATAAGAAATGCAATAATTGAAACTGAAAAATAATTAATTTATAGTGA 2758
QY 193 TCAAAGTAGTTATCTAGTAGTGTAATTAATA--ATCGATGCGCGATTCAGAAATGGG 250
DB 2757 TCACAACAAGTTTGAATTAATTTAAACTATATTATAATTAATTAATTAATTAATTAAT 2698
QY 251 ACAACAATGAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
DB 2697 CTATAAATAAATAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 2638
QY 311 TCTTTTCTGACTATTGAGGGCAAA 336
DB 2637 TTTTATATATCAATTAAGCAGACAA 2612

RESULT 10
US-10-239-676-100/c
Sequence 100, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 100

LENGTH: 17419
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-100
Query Match
Best Local Similarity 4.5%; Score 49.2; DB 9; Length 17419;
Matches 173; Conservative 0; Mismatches 188; Indels 1; Gaps 1;
QY 660 ATCTAATGACTTTTTTTTTTCTACCCGCTGGATGAAAGTTATATAGTACTATAGCCAGAG 719
DB 11067 ATCAACAATCTTACTTACTTAAATCTTTTCAACAATATATCTACTTAACTATATTTCTACT 11008
QY 720 ACAATTGATTATAGATATATCCATTAATCCATGATATTTATGATATAATAGCTGTAA 779
DB 11007 TTACTTCTTATTTTATTTTACTTATTAATCAAACTTACTTCTCTTTCCAAATACTTTTAT 10948
QY 780 CTATTTTCAGATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTTAAAGATTTTAATAATAA 839
DB 10947 TTTAATAATAAAATCTTCTTTTAAAAAACTAATTAATAATAATAATAATAATAATAA 10888
QY 840 AAGTATTAAGGAGCATAACGAGCGACAAAGTAATGAACGAGAGAAACAAAGCCA 899
DB 10887 TAATAATAACAATAATTAATAACAATAATAATATATCTAAACCAATCTAAATAATAA 10828
QY 900 TGAAGCTCATTGGTTAGTTTAAAGCTTAATAGAGATTTTATTAATTTTAAATGACGATG 959
DB 10827 CTCA-CATAATATATCTTTCGTTAATAAACAATAATAATATATTTATCAACTCCATA 10769
QY 960 ATACAATTAATTTTCTGACTTCTTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 1019
DB 10768 AAAACAATATATATATATATCTTATTTTATCTATATCTTAAACCTTAAACCTTAA 10709
QY 1020 TC 1021
DB 10708 TC 10707

RESULT 11
US-10-224-562-3/c
Sequence 3, Application US/10224562
Publication No. US2003002229A1
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL001098DIV
CURRENT APPLICATION NUMBER: US/10/224,562
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 53332
TYPE: DNA
ORGANISM: Homo sapiens
US-10-224-562-3

Query Match
Best Local Similarity 4.4%; Score 48.6; DB 9; Length 53332;
Matches 129; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
QY 47 TCCAAAAAAATTTTGTAAACAGAACTTCCAAATTTTTTTTTTATGGAACAGAAAT 106
DB 31671 TCAAAAAAAATTAATATATATATATAAAAAATAATAATAATAATAATAATAATAT 31612
QY 107 AACAGATAGAAACTATTTTGTGGAATGGAAGTACTATATATACATTAGCAAAATTT 166
DB 31611 AAATATAAAATATATATATATAATAATAATAATAATAATAATAATAATAATAATAT 31552
QY 167 AAAAAATTAATAGCCTATACGGCTCAAAAGTATGTTATCTAGTAGTGTAAATAATA 226


```
Db      88  AAAAAAAAAACANNAGANNANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 147
OY      61  TGTACAGAACTCCAAATTTTTTTTTTATGAGACAGAAATTAACGATGGAAC 120
Db      148  ANAAAAAANCANNAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 207
OY      121  TATTTGTTGTAAGTGAAGTACTAATATACATTAGCAATTTTAAAAAATATATTA 180
Db      208  AANNANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 267
OY      181  GCCTATACGCCCTAAAGTATGTTATCTAGTAGTAAATTAATGATGTCGAT 240
Db      268  NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 327
OY      241  CAGATTGGACAAACATGAAAGCAATTAATAATTAATTAATTAATTAATTAAT 300
Db      328  AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAATAAAA 387
OY      301  TGAGTAAATGTGTTCTGACTATGAGGGCAAAAAAAGACATGCCAAAGCTTA 358
Db      388  TNAAAAAATATNTAANAANAATATNTAANAATAAAAAANAAAAAANAAAAAATA 445
```

RESULT 15

```
US-10-239-676-51/C
; Sequence 51, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51
```

Query Match 4.4%; Score 48; DB 9; Length 9539;
Best Local Similarity 47.9%; Pred. No. 7.9;

Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
OY      27  AATAAATGATGAGAAAGTCCAAAAAATTTTGTAAACAGAACTCCAAATTTT 86
Db      6023  ATTAAAAATATATTAATCAAAATCTTAAATCATAAAAATCTAAAAAATTCAAA 5964
OY      87  TTTTGTGGAAGAAATTAACAGATGAAACTATTTTGTGGAATGGAAGTAGTA 146
Db      5963  AATTTCTAAAAAATTAATTAATAATCAAAAAATCTATTTAAACAAACAAATA 5904
OY      147  ATATACATTAGCAATTTTAAAAAATTAATTAAGCCTATACGCGCTCAAGTAGTAT 206
Db      5903  TAACATAATAAAAAATTTAAAAACCAAAATTAATAATAAAAAATTTAAAAA 5844
OY      207  CTAGTAGTGTAATTAATTAATGATGCGATTCAGAAATTTGGACAAATGAAAAACGG 266
```

```
Db      5843  CTAATAAAATTCAAAAAACAATAAATATATTTTAAACTAAATACATTTAAAAA 5784
OY      267  AATTAAATTTTAACTTTAAATTAATTAATTAATTTGAGTAATGTGTT 314
Db      5783  ACAAAATATAAAATTTTATACGACAAACAAATTTTAAAAAATTAATTT 5736
```

Search completed: June 11, 2003, 08:07:37
Job time: 159.028 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 2645.14 Seconds
(without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaataattaatat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 segs, 12571243825 residues 49582208

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Database: Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/1/pna/US06000.COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06001.COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US06002.COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US06003.COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US06004.COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US06005.COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US06006.COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US06007.COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US06008.COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US06009.COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US06010.COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US06011.COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US06012.COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US06013.COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US06014.COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US06015.COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US06016.COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US06017.COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US06018.COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US06019.COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US06020.COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US06021.COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US06022.COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US06023.COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US06024.COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US06025.COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US06026.COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US06027.COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US06028.COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US06029.COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US06030.COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US06031.COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US06032.COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US06033.COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US06034.COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US06035.COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US06036.COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US06037.COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US06038.COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US06039.COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US06040.COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US06041.COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US06042.COMB.seq.*

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	6888	1	Sequence 1, Appl1
2	1101	100.0	6888	1	Sequence 1, Appl1
3	1090	99.0	84196	27	Sequence 94, Appl1
4	1090	99.0	84196	31	Sequence 1074, Ap
5	960.4	87.2	1629	18	Sequence 2050, Ap
6	125	11.4	826	1	Sequence 3486, Ap
7	66.6	6.0	469	33	Sequence 40543, A
8	66.6	6.0	469	33	Sequence 40543, A
9	65	5.9	450	17	Sequence 84816, A
10	65	5.9	528	25	Sequence 15000, A
11	65	5.9	528	25	Sequence 121875, A
12	61	5.5	553	22	Sequence 121875, A
13	60.4	5.5	507	33	Sequence 29073, A
14	60.4	5.5	507	33	Sequence 91952, A
15	59.8	5.4	358	27	Sequence 135967, A
16	59.8	5.4	358	27	Sequence 340019, A
17	59.8	5.4	358	27	Sequence 340019, A
18	59.6	5.4	422	33	Sequence 6397, Ap
19	59.6	5.4	547	33	Sequence 30413, Ap
20	59.6	5.4	547	33	Sequence 1402, Ap
21	5.4	5.4	664	29	Sequence 1402, Ap
					Sequence 739, App

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

22 59 5 4 664 31 US-09-803-110-739 Sequence 739, App
23 58.8 5.3 2000 33 US-09-887-272A-5263 Sequence 5263, Ap
24 58.4 5.3 447 25 US-09-654-617-341141 Sequence 341141,
25 58.4 5.3 447 27 US-09-684-016-341141 Sequence 341141,
26 58.4 5.3 447 28 US-09-145-485-9030 Sequence 9030, Ap
27 58.4 5.3 720 18 US-09-404-520-12953 Sequence 12953, A
28 58.4 5.3 805 56 US-09-126-265-127 Sequence 127, App
29 58.4 5.3 961 56 US-09-126-265-3039 Sequence 3039, Ap
30 57.6 5.2 513 25 US-09-654-617-181496 Sequence 181496,
31 57.2 5.2 513 27 US-09-684-016-181496 Sequence 181496,
32 57.2 5.2 676 33 US-09-865-419A-17337 Sequence 17337, A
33 57.2 5.2 676 33 US-09-865-419A-17337 Sequence 17337, A
34 56.8 5.2 485 23 US-09-619-643-59 Sequence 59, App
35 56.8 5.2 485 23 US-09-654-617-334306 Sequence 334306,
36 56.8 5.2 485 27 US-09-684-016-334306 Sequence 334306,
37 56.8 5.2 485 28 US-09-146-907-71 Sequence 71, Appl
38 56.6 5.1 407 24 US-09-637-086A-28965 Sequence 28965, A
39 56.6 5.1 407 24 US-09-637-086D-28965 Sequence 28965, A
40 56.6 5.1 6107 42 US-10-221-613-304 Sequence 304, App
41 56.4 5.1 469 25 US-09-654-617-86668 Sequence 86668, A
42 56.4 5.1 469 27 US-09-684-016-86668 Sequence 86668, A
43 56.2 5.1 534 22 US-09-565-306-13731 Sequence 13731, A
44 56 5.1 501 33 US-09-873-402A-2082 Sequence 2082, Ap
45 56 5.1 501 64 US-09-209-830-2082 Sequence 2082, Ap

```

ALIGNMENTS

```

RESULT 1
PCT-US00-03820-1
; Sequence 1, Application PC/TUS0003820
; GENERAL INFORMATION:
; APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001.40
; CURRENT APPLICATION NUMBER: PCT/US00/03820
; CURRENT FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: 60/119,657
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/119,658
; EARLIER FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; PCT-US00-03820-1

```

```

Query Match 100.0%; Score 1101; DB 1; Length 6888;
Best Local Similarity 100.0%; Pred. No. 2.4e-189;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AATCTACAATTTTATATATTTAGTCAATPACAAATGCATGAAGTCCAAAAAATTT 60
DB 2102 AATCTACAATTTTATATATTTAGTCAATPACAAATGCATGAAGTCCAAAAAATTT 2161
QY 61 TGTTAACAGAACTTCCAAATTTTTTTTTTATGAAACAAGAAATTAACGATAGAAAC 120
DB 2162 TGTTAACAGAACTTCCAAATTTTTTTTTTATGAAACAAGAAATTAACGATAGAAAC 2221
QY 121 TATTTTGTGTGGAATGGAATAGTATATATACATTAAGCAAAATTTAAAAAATTAATA 180
DB 2222 TATTTTGTGTGGAATGGAATAGTATATATACATTAAGCAAAATTTAAAAAATTAATA 2281
QY 181 GCCTATACGCGCTCAAGATGTATCTAGTACGTGAATTAATGATGCTGTGGCAT 240
DB 2282 GCCTATACGCGCTCAAGATGTATCTAGTACGTGAATTAATGATGCTGTGGCAT 2341
QY 241 CAGAAATGGGACAAACATGAAGAAAGCAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 2342 CAGAAATGGGACAAACATGAAGAAAGCAATTAATTAATTAATTAATTAATTAATTAAT 2401

```

```

QY 301 TGAATAATGTTTCTGACTATGTAGGGGCAAAAAAAGACAAATGCCAAAGCTACG 360
DB 2402 TGAATAATGTTTCTGACTATGTAGGGGCAAAAAAAGACAAATGCCAAAGCTACG 2461
QY 361 GGTTCAGTCCGAGTCCGTAATATCTATATCTGTCTTTGACCCGACGCTCGT 420
DB 2462 GGTTCAGTCCGAGTCCGTAATATCTATATCTGTCTTTGACCCGACGCTCGT 2521
QY 421 AGGGGCTCTTCTGACATTTTCACGTCTTCACTCCCTACTCGAGACCCACTTTCCAT 480
DB 2522 AGGGGCTCTTCTGACATTTTCACGTCTTCACTCCCTACTCGAGACCCACTTTCCAT 2581
QY 481 ATCCTAAGGTAATTTTGGAAATCCCAATTTTAAACGATGAGACCGTACCGACTTCT 540
DB 2582 ATCCTAAGGTAATTTTGGAAATCCCAATTTTAAACGATGAGACCGTACCGACTTCT 2641
QY 541 GGGATTCGTCGGGCAATTTATCAAAATTTATGCAAGATGGGTTTATTAATTA 600
DB 2642 GGGATTCGTCGGGCAATTTATCAAAATTTATGCAAGATGGGTTTATTAATTA 2701
QY 601 ACTCACAACTTGATCAGATTAATAATTTTATCAACACTTTTACGATGATTCGTACGATCTA 660
DB 2702 ACTCACAACTTGATCAGATTAATAATTTTATCAACACTTTTACGATGATTCGTACGATCTA 2761
QY 661 TCTAATGACTTTTCTTCTACACGCTGAGTAAGTATAGTACTATTAGCCAGAGA 720
DB 2762 TCTAATGACTTTTCTTCTACACGCTGAGTAAGTATAGTACTATTAGCCAGAGA 2821
QY 721 CAATGATTAATAGTATATATCCATTAATCCCATGATTTTATGATTAATTAATTAATTAAT 780
DB 2822 CAATGATTAATAGTATATATCCATTAATCCCATGATTTTATGATTAATTAATTAATTAAT 2881
QY 781 TATTTCAAGACGCGAGCTTCTGCACTTTGTTTATTAATTAAGATTTAATAATTA 840
DB 2882 TATTTCAAGACGCGAGCTTCTGCACTTTGTTTATTAATTAAGATTTAATAATTA 2941
QY 841 AGTATTAAGAGACATPACAGAGCAACAAAGTATGACAGCGAGAAACAAAGCAT 900
DB 2942 AGTATTAAGAGACATPACAGAGCAACAAAGTATGACAGCGAGAAACAAAGCAT 3001
QY 901 GAAGCTCATTTGGTATGTTTAAGCTTAATTAAGAAAGATTTTATTAATTTAATGACATGA 960
DB 3002 GAAGCTCATTTGGTATGTTTAAGCTTAATTAAGAAAGATTTTATTAATTTAATGACATGA 3061
QY 961 TAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGAACTCCCTTTT 1020
DB 3062 TAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGAACTCCCTTTT 3121
QY 1021 CAGTAGAAGTCCGATTTCCCAATCTTAAAGACAAAGCCATTGAAAGAGAAAGTAGTAG 1080
DB 3122 CAGTAGAAGTCCGATTTCCCAATCTTAAAGACAAAGCCATTGAAAGAGAAAGTAGTAG 3181
QY 1081 AGAGAGAGAGAACTAGCTCC 1101
DB 3182 AGAGAGAGAGAACTAGCTCC 3202

```

```

RESULT 2
US-09-502-426-1
; Sequence 1, Application US/09502426
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426
; CURRENT FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: 60/119,657
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/119,658
; EARLIER FILING DATE: 1999-02-11

```

NUMBER OF SEQ ID NOS: 18
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 1
 LENGTH: 6888
 TYPE: DNA
 ORGANISM: Arabidopsis sp.
 US-09-502-426-1

Query Match 100.0%; Score 1101; DB 19; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 2,4e-189;
 Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AATCTCAAAATTTATTTATTTAGTCAATTAACATGATAGCAAGTCCAAAAAATTT 60
 2102 AATCTCAAAATTTATTTATTTAGTCAATTAACATGATAGCAAGTCCAAAAAATTT 2161
 61 TGTTAACAGAACTCCAAATTTTTTTTTATGCAACAGAAATTAACAGATGAAAC 120
 2162 TGTTAACAGAACTCCAAATTTTTTTTTATGCAACAGAAATTAACAGATGAAAC 2221
 121 TATTTGTTGGATGAGTAAATATATCATATGATAGCAATTTTAAAAATTTATTA 180
 2222 TATTTGTTGGATGAGTAAATATATCATATGATAGCAATTTTAAAAATTTATTA 2281
 181 GCTTATAGCGCTCAAGATATGTATCTAGTGTATTAATTAATGATGTCGAT 240
 2282 GCTTATAGCGCTCAAGATATGTATCTAGTGTATTAATTAATGATGTCGAT 2341
 241 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 300
 2342 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 2401
 301 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCGCAAGTCTAC 360
 2402 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCGCAAGTCTAC 2461
 361 GGTGACATGTCAGTGGTATATATCTGATTAATGATGTCGTCGTCGTCGTCG 420
 2462 GGTGACATGTCAGTGGTATATATCTGATTAATGATGTCGTCGTCGTCGTCG 2521
 421 AGGGTCTCTGACATTTTCACTGTTTACCCCTACCTGTCGTCGTCGTCGTCGTC 480
 2522 AGGGTCTCTGACATTTTCACTGTTTACCCCTACCTGTCGTCGTCGTCGTCGTC 2581
 481 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACGATGAGCGTACCGGATCTCT 540
 2582 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACGATGAGCGTACCGGATCTCT 2641
 541 GGGATTCGTCGAGCATTTATCAAAATTTATTAGCAGCAATGGGTTTATTAATTA 600
 2642 GGGATTCGTCGAGCATTTATCAAAATTTATTAGCAGCAATGGGTTTATTAATTA 2701
 601 ACTCACAATCTATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
 2702 ACTCACAATCTATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2761
 661 TCTAATGACATTTTCTACCAAGGTCGATGAAAGTATAGTACTTTTGGCAGAGA 720
 2762 TCTAATGACATTTTCTACCAAGGTCGATGAAAGTATAGTACTTTTGGCAGAGA 2821
 721 CAATGATTAATGATATATCATTAATCATGATATTAATTAATTAATTAATTAAT 780
 2822 CAATGATTAATGATATATCATTAATCATGATATTAATTAATTAATTAATTAAT 2881
 781 TATTTACAGATCGCAGCTTTTGCACATTTTGTATTAATTAATTAATTAATTA 840
 2882 TATTTACAGATCGCAGCTTTTGCACATTTTGTATTAATTAATTAATTAATTA 2941
 841 ACTATTAAGAGCATTAAGAGCAAAAGTATTAATTAATTAATTAATTAATTAAT 900
 2942 ACTATTAAGAGCATTAAGAGCAAAAGTATTAATTAATTAATTAATTAATTAAT 3001
 901 GAAGTCATGTTGTTAGTTTAACTTAATTAAGAAATTTTAAATTTTAAATGAGCAT 960

|||||
 Db 3002 GAAGTCATGTTGTTAGTTTAACTTAATTAAGAAATTTTAAATTTTAAATGAGCAT 3061
 961 TAACATTAATTTTCTGACATTTTAAACCCCTTTTAAACAGAGATCCCTTTT 1020
 3062 TAACATTAATTTTCTGACATTTTAAACCCCTTTTAAACAGAGATCCCTTTT 3121
 1021 CAGTAGAGTCCGATTTCCCAATTAAGAGCAAAAGTATTAATTAATTAATTAAT 1080
 3122 CAGTAGAGTCCGATTTCCCAATTAAGAGCAAAAGTATTAATTAATTAATTAAT 3181
 1081 AGAGAGAGAACTAGCTCC 1101
 3182 AGAGAGAGAACTAGCTCC 3202

RESULT 3
 US-09-692-412-94/C
 Sequence 94, Application US/09692412
 GENERAL INFORMATION:
 APPLICANT: Bush, David F.
 APPLICANT: Rounsley, Steven D.
 APPLICANT: Wiegand, Roger C.
 TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
 FILE REFERENCE: 38-10(15493)C
 CURRENT APPLICATION NUMBER: US/09/692,412
 NUMBER OF SEQ ID NOS: 124
 SEQ ID NO 94
 LENGTH: 84196
 TYPE: DNA
 ORGANISM: Arabidopsis thaliana
 US-09-692-412-94

Query Match 99.0%; Score 1090; DB 27; Length 84196;
 Best Local Similarity 99.9%; Pred. No. 3.6e-187;
 Matches 1101; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 AATCTCAAAATTTATTTATTTAGTCAATTAACATGATAGCAAGTCCAAAAAATTT 60
 30101 AATCTCAAAATTTATTTATTTAGTCAATTAACATGATAGCAAGTCCAAAAAATTT 30042
 61 TGTTAACAGAACTCCAAATTTTTTTTTATGCAACAGAAATTAACAGATGAAAC 120
 30041 TGTTAACAGAACTCCAAATTTTTTTTTATGCAACAGAAATTAACAGATGAAAC 29982
 121 TATTTGTTGGATGAGTAAATATATCATATGATAGCAATTTTAAAAATTTATTA 180
 29981 TATTTGTTGGATGAGTAAATATATCATATGATAGCAATTTTAAAAATTTATTA 29922
 181 GCTTATAGCGCTCAAGATATGTATCTAGTGTATTAATTAATTAATTAATTAAT 240
 29921 GCTTATAGCGCTCAAGATATGTATCTAGTGTATTAATTAATTAATTAATTAAT 29862
 241 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 300
 29861 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 29802
 301 TGAGTAATATGTTTCTGACATTTTGGAGGCAAAAAAGCAATGCGCAAGTCTAC 360
 29801 TGAGTAATATGTTTCTGACATTTTGGAGGCAAAAAAGCAATGCGCAAGTCTAC 29742
 361 GGTGACATGTCAGTGGTATATATCTGATTAATGATGTCGTCGTCGTCGTCG 420
 29741 GGTGACATGTCAGTGGTATATATCTGATTAATGATGTCGTCGTCGTCGTCG 29682
 421 AGGGTCTCTGACATTTTCACTGTTTACCCCTACCTGTCGTCGTCGTCGTCGTC 480
 29681 AGGGTCTCTGACATTTTCACTGTTTACCCCTACCTGTCGTCGTCGTCGTCGTC 29622
 481 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACGATGAGCGTACCGGATCTCT 540
 29621 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACGATGAGCGTACCGGATCTCT 29562

```

QY      541  GGGATTCGCGGAGCATTATTCACAAAATTTATACAGAAAGGGTTATTAATTAATAA 600
Db      29561 GGGATTCGCGGAGCATTATTCACAAAATTTATACAGAAAGGGTTATTAATTAATAA 29562
QY      601  ACTCAACAACCTTGATCAGATTAATAATTTTCATPAACACTTTTACGATGGATTCGACATCTA 660
Db      29501  ACTCAACAACCTTGATCAGATTAATAATTTTCATPAACACTTTTACGATGGATTCGACATCTA 29442
QY      661  TCTAATGACTTTTTTTTTTTCTACACGGGTGATGAAGTTTATAGTACTATTAAGCCAGAGA 720
Db      29441  TCTAATGACTTTTTTTTTTTCTACACGGGTGATGAAGTTTATAGTACTATTAAGCCAGAGA 29382
QY      721  CAATTGATTAATAGATATATATCCATTATATCCATGATATTTATGATATATAATAGCTGTTAAAC 780
Db      29381  CAATTGATTAATAGATATATATCCATTATATCCATGATATTTATGATATATAATAGCTGTTAAAC 29322
QY      781  TATTTGACGATCGCGACTTTCTGCAACTTTTGTTTTAAATTTAAGAGTTTAAATTAATTA 840
Db      29321  TATTTGACGATCGCGACTTTCTGCAACTTTTGTTTTAAATTTAAGAGTTTAAATTAATTA 29262
QY      841  AGATATTAAGAAGAGCATATACGAGCAACAAAGTATATACACGAGAGAACAAAGCCAT 900
Db      29261  AGATATTAAGAAGAGCATATACGAGCAACAAAGTATATACACGAGAGAACAAAGCCAT 29202
QY      901  GAAGCTCATTTGGTTA-GTTTAAGCTTAATPAAGAAAGATTTTATTAATTTTATGACGATG 959
Db      29201  GAAGCTCATTTGGTTAAGTTTAAGCTTAATPAAGAAAGATTTTATTAATTTTATGACGATG 29142
QY      960  ATACCAATTAATTTCTGACTCTGTTAAACCCTCTTACAAACACAGACTCCCTTTT 1019
Db      29141  ATACCAATTAATTTCTGACTCTGTTAAACCCTCTTACAAACACAGACTCCCTTTT 29082
QY      1020  TCAGTAGAAGTCCGATTCCTTAATGAACCAAGCCATTAGAAGAAGAAAGTAGATGA 1079
Db      29081  TCAGTAGAAGTCCGATTCCTTAATGAACCAAGCCATTAGAAGAAGAAAGTAGATGA 29022
QY      1080  GAGAGAGAGAGAAACTAGCTCC 1101
Db      29021  GAGAGAGAGAGAAACTAGCTCC 29000

RESULT 4
US-09-803-736-1074/c
; Sequence 1074, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 1074
; LENGTH: 84196
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-1074

```

Query Match	99.0%	Score 1090	DB 31	Length 84156
Best Local Similarity	99.9%	Pred. No. 3	6e-187	
Matches 1101: Conservative	0	Mismatches	0	Indels 1: Gaps 1:
OY	1	AATCTACAAATTTTAATATTGTCATATACATCATGACATGAGAAAGTTCACAAAAAATTTT	60	
Db	30101	AATCTACAAATTTTAATATTGTCATATACATCATGAGAAAGTTCACAAAAAATTTT	3004 22	

OY	61	TGTTAACGAAACCTCCAAATTTTTTTTTTTTNGGAAACGAAATACAGATAGCAAAAC	120
Db	30041	TGTTAACGAAACCTCCAAATTTTTTTTTTTTNGGAAACGAAATACAGATAGCAAAAC	29982
OY	121	TATTTTGTGTGSAATGGAAGTAGATAATATACATTAAGCAAAATTTTAAAAATTAATTA	180
Db	29981	TATTTTGTGTGSAATGGAAGTAGATAATATACATTAAGCAAAATTTTAAAAATTAATTA	29922
OY	181	GCCATACGGCCCTCAAGTAGTAGTACTAGTCAATTAATTAATGATGCTGGAT	240
Db	29921	GCCATACGGCCCTCAAGTAGTAGTACTAGTCAATTAATTAATGATGCTGGAT	29862
OY	241	CAGAAATGGGCAACAATGAAAAAGGAAATTAATTAATTAATTAATTAATTAATTAAT	300
Db	29861	CAGAAATGGGCAACAATGAAAAAGGAAATTAATTAATTAATTAATTAATTAATTAAT	29802
OY	301	TGAGTAAATGTGTTTTCTGACTATTTAGGGGCAAAAAAGACATGCCAAAGCTACG	360
Db	29801	TGAGTAAATGTGTTTTCTGACTATTTAGGGGCAAAAAAGACATGCCAAAGCTACG	29742
OY	361	GGTTTGAAGTCGCCAGTTCCGGTAATTAATTAATTAATTAATTAATTAATTAATTAAT	420
Db	29741	GGTTTGAAGTCGCCAGTTCCGGTAATTAATTAATTAATTAATTAATTAATTAATTAAT	29682
OY	421	AGGGGTCCTTGTGACATTTTCACTGTTCTACCCCTACTCGTAGGCCCAACCTTTCCAT	480
Db	29681	AGGGGTCCTTGTGACATTTTCACTGTTCTACCCCTACTCGTAGGCCCAACCTTTCCAT	29622
OY	481	ATCCTAAGGGGAATTTTGGAAATCCCAATTAACGATTAAGACCTGACCGGACTCCT	540
Db	29621	ATCCTAAGGGGAATTTTGGAAATCCCAATTAACGATTAAGACCTGACCGGACTCCT	29562
OY	541	GGGATTCCTGTGAGACATTTATCAAAAAATTATTAGCAGCAATGGGTTTATTAATTAAA	600
Db	29561	GGGATTCCTGTGAGACATTTATCAAAAAATTATTAGCAGCAATGGGTTTATTAATTAAA	29502
OY	601	ACTACAACTTGATGACATATAAATTTTCATTAACCTTTACGATGATTCGACATCTA	660
Db	29501	ACTACAACTTGATGACATATAAATTTTCATTAACCTTTACGATGATTCGACATCTA	29442
OY	661	TCTAATACCTTTTTTTTTTCTACACGGTGGATGAAGTATATAGTACTTTAGCCAGAGA	720
Db	29441	TCTAATACCTTTTTTTTTTCTACACGGTGGATGAAGTATATAGTACTTTAGCCAGAGA	29382
OY	721	CAATTGATTAATGATATATTCATTAATTCATGATATTTATGATTAATTAATAGCTGTTAAC	780
Db	29381	CAATTGATTAATGATATATTCATTAATTCATGATATTTATGATTAATTAATAGCTGTTAAC	29322
OY	781	TATTTGAGATCGAGCTTCTGCAACTTTGTTTTATTTAATTAAGGTTTATTAATTAATAA	840
Db	29321	TATTTGAGATCGAGCTTCTGCAACTTTGTTTTATTTAATTAAGGTTTATTAATTAATAA	29262
OY	841	AGTATTAATAAGGACATTAACGAGCAACAAAGTAAATGAACAACGGAGAAACAAAGCCAT	900
Db	29261	AGTATTAATAAGGACATTAACGAGCAACAAAGTAAATGAACAACGGAGAAACAAAGCCAT	29202
OY	901	GAAGCTCATTTGGTTA	959
Db	29201	GAAGCTCATTTGGTTA	29142
OY	960	ATAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGACAGCTCCCTTTT	1019
Db	29141	ATAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGACAGCTCCCTTTT	29082
OY	1020	TCAGTAGAAGTCGGATTTCCCAATCTTTAAAGCAAAAGCATTTGAAAGAGAAAGTAGTGA	1079
Db	29081	TCAGTAGAAGTCGGATTTCCCAATCTTTAAAGCAAAAGCATTTGAAAGAGAAAGTAGTGA	29022
OY	1080	GAGAGAGAGAGAAACTAGCTCC 1101	
Db	29021	GAGAGAGAGAGAAACTAGCTCC 29000	

RESULT 5
 US-09-451-320-2050
 ; Sequence 2050, Application US/09451320
 ; GENERAL INFORMATION:
 ; APPLICANT: ALEXANDROV, Nikolai
 ; APPLICANT: CHEN, Xianfeng
 ; APPLICANT: SUBRAMANIAN, Gopalakrishnan
 ; APPLICANT: ZHENG, Liansheng
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 42-59, 62-66, 68, 81
 ; FILE REFERENCE: 2750-0663P
 ; CURRENT APPLICATION NUMBER: US/09/451,320
 ; NUMBER OF SEQ ID NOS: 6998
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2050
 ; LENGTH: 1699
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; OTHER INFORMATION: any n or Xaa = unknown
 US-09-451-320-2050

Query Match
 Best Local Similarity 87.2%; Score 960.4; DB 18; Length 1699;
 Matches 983; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

DB 107 AACAGATGAGAAACCTATTTTGTGGAATGGAAGTAAATATCATTAAGCAATTTT 166
 DB 1 AACAGATGAGAAACCTATTTTGTGGAATGGAAGTAAATATCATTAAGCAATTTT 60
 QY 167 AAAAAATTTATTAAGCCTATACGCGCTCAAAAGTATGTAATCTAGTAGGTAAATTA 226
 DB 61 AAAAAATTTATTAAGCCTATACGCGCTCAAAAGTATGTAATCTAGTAGGTAAATTA 120
 QY 227 TGCATGTGCGATTTAGAAATGGAGCAACAATGAATGAATGAATGAATGAATGA 286
 DB 121 TGCATGTGCGATTTAGAAATGGAGCAACAATGAATGAATGAATGAATGAATGA 180
 QY 287 AATAATATAAATTTAGTAATGTCTTCTGACATTTTCTAGGCGCAAAAAAGCAAT 346
 DB 181 AATAATATAAATTTAGTAATGTCTTCTGACATTTTCTAGGCGCAAAAAAGCAAT 240
 QY 347 GCCAAAGTCTAGCGGTTTGTGACGTCCAGTTCGTAATATCATTAATCATCTGTCTGA 406
 DB 241 GCCAAAGTCTAGCGGTTTGTGACGTCCAGTTCGTAATATCATTAATCATCTGTCTGA 300
 QY 407 CCGAGCGCTGTAGGGGTCTTCTGACATTTTCTAGGCGCAAAAAAGCAAT 466
 DB 301 CCGAGCGCTGTAGGGGTCTTCTGACATTTTCTAGGCGCAAAAAAGCAAT 360
 QY 467 CACCGCTTTCCATATCTTAAGGTAATTTGAAATCCCAATTAACCGATGAGAGC 526
 DB 361 CACCGCTTTCCATATCTTAAGGTAATTTGAAATCCCAATTAACCGATGAGAGC 420
 QY 527 GTACCGGATCTCTGGATCTGCTGAGCAATTTATCAAAATTTATTAAGCAAGAGGT 586
 DB 421 GTACCGGATCTCTGGATCTGCTGAGCAATTTATCAAAATTTATTAAGCAAGAGGT 480
 QY 587 TATTAATTTAAATCTCACAACCTGATGATTAATTTCAATTAACCTTTAGCATGG 646
 DB 481 TATTAATTTAAATCTCACAACCTGATGATTAATTTCAATTAACCTTTAGCATGG 540
 QY 647 ATTGATGATCTCTTAATGATCTTTTCTTCTACGAGCGGTGAAGATTAATGA 706
 DB 541 ATTGATGATCTCTTAATGATCTTTTCTTCTACGAGCGGTGAAGATTAATGA 600
 QY 707 CTATTAGCCAGAGCAATTTATTAATTAATCATTAATCATTAATTAATTAATGA 766
 DB 601 CTATTAGCCAGAGCAATTTATTAATTAATCATTAATCATTAATTAATTAATGA 660
 QY 767 AATAGCTGTAATCTTTTACAGATGAGCTTCTGCAACTTTGTTTAAATTAAGA 826
 DB 826 AATAGCTGTAATCTTTTACAGATGAGCTTCTGCAACTTTGTTTAAATTAAGA

DB 661 AATAGCTGTAATCTTTTACAGATGAGCTTCTGCAACTTTGTTTAAATTAAGA 720
 QY 827 GTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 886
 DB 721 GTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
 QY 887 GAAACAAAAGCCATGAGCTATTTGTTA-GTTTAAGCTTAATTAAGATTTTAA 945
 DB 781 GAAACAAAAGCCATGAGCTATTTGTTAAGCTTAATTAAGATTTTAA 840
 QY 946 TTTTAATGAC-GATGATTAATTAATTTTCTGACCTTTTAAACCCCTTACAA 1004
 DB 841 TTTTAATGACGATGATTAATTTTCTGACCTTTTAAACCCCTTACAA 900
 QY 1005 CAGAGCTCCCTTTTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
 DB 901 CAGAGCTCCCTTTTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 1065 AGAGAAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 DB 961 AGAGAAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

RESULT 6
 PCT-US99-22853B-3486
 ; Sequence 3486, Application PC/TUS9922853B
 ; GENERAL INFORMATION:
 ; APPLICANT: Ceres, Inc.
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
 ; FILE REFERENCE: Polypeptides Encoded Thereby
 ; CURRENT APPLICATION NUMBER: PCT/US99/22853B
 ; NUMBER OF SEQ ID NOS: 3938
 ; SOFTWARE: MS Word 97
 ; SEQ ID NO 3486
 ; LENGTH: 826
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; OTHER INFORMATION: LOCATION 1..826, Ceres Seq. ID 1592085
 ; LOCATION: (1)..(826)
 ; OTHER INFORMATION: any n = any nucleic acid, unknown or other
 PCT-US99-22853B-3486

Query Match
 Best Local Similarity 11.4%; Score 125; DB 1; Length 826;
 Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 977 TGACCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTTACAGTAAAGTCCGATT 1036
 DB 1 TGACCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTTACAGTAAAGTCCGATT 60
 QY 1037 CCAATCTTAAAGCAAAAGCCATTAAGAAAGAAAGTGAAGAGAGAGAGAACTA 1096
 DB 61 CCAATCTTAAAGCAAAAGCCATTAAGAAAGAAAGTGAAGAGAGAGAGAACTA 120
 QY 1097 GCTCC 1101
 DB 121 GCTCC 125

RESULT 7
 US-09-865-439A-40543/C
 ; Sequence 40543, Application US/09865439A
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; APPLICANT: Hardeman, Kristine J.
 ; APPLICANT: La Rosa, Thomas J.
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 ; FILE REFERENCE: 38-21(51936)B


```

? CURRENT APPLICATION NUMBER: US/09/865,439A
? CURRENT FILING DATE: 2001-05-29
? PRIOR APPLICATION NUMBER: US 60/207,458
? PRIOR FILING DATE: 2000-05-30
? NUMBER OF SEQ ID NOS: 119126
? SEQ ID NO 40543
? LENGTH: 469
? TYPE: DNA
? ORGANISM: Zea mays
? FEATURE:
? OTHER INFORMATION: Clone ID: LBB335-005-P1-K1-G2
US-09-865-439A-40543
```

Query Match	6.0%	Score	66.6	DB	33	Length	469
Best Local Similarity	50.1%	Pred. No.	0.031				
Matches	193	Conservative	0	Mismatches	189	Indels	3
						Gaps	1

OY		9	AATATTAAATTTTGGTCATAATACAGCATGGAAGCTGCACAAAAAATTTTGTAACA	68
Dd		463	AAAATTAATTTATATTATTTATTTAAACCTAATAATTAATTAATTAATTAATTAATTA	404
OY		69	GAAACTCCCAATTTTTTTTTTTATGGAACAAGAATAATACAGATGGAACACTATTGTG	128
Dd		403	TAAAGATAAAAATTTTATTAATAAAATTAATAAAATGTTATATATATTAATAAAATTAAT	344
OY		129	TGTGGAATGCAAGTAGTAATATACCTTAAGCAAATTTTAAAAAATTTATATTAAGCCTATAC	188
Dd		343	AATGAAAAAAAAAATATATATATAAAATAAATAATATATGTTAAAAATTTATATATAAATAAAA	284
OY		189	GGCGTCAGAGTAGTATCTATCTAGTAGCTGAATTA--ATAATGACATGGGCCGATTCAGAA	245
Dd		283	ATATATTAATAAATATTTTAAAGTATATATATAGTAATATATATATATTAAGATTTATATA	224
OY		246	TTGGACACACATGAAAAACGCAATTAATAATATTAACCTTAATAATAATAATAAATTTGAGT	305
Dd		223	TTTAAAAAATAAAATGTAATATATTTTATTAATAAAATTTTAAATTTAACCTATTTATATAAAA	164
OY		306	AAATGTGTTTCTGACTATTTGAGGGCACAAAAAAGACATGCCAAAAGCTTACGGGTTT	365
Dd		163	AATGTGTTTTTAATTTAAAAAATAAATAAATTTATTAATTAATATATATATATTTT	104
OY		366	GACTGTCAGTTGGGTAAATCTA	390
Dd		103	TATTTTATTTAATTAATAATTTATA	79

RESULT 8
US-60-207-458-84816/c
; Sequence 84816, Application US/60207458
; Original Truncation

APPLICANT: Abad, Mark S.
 APPLICANT: Conner, Timothy W.
 APPLICANT: Delkman, Jill
 APPLICANT: Hardeman, Kristine J.
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Laigudi, Raghunath V.
 APPLICANT: Ruan, Yijun G.
 APPLICANT: Ruff, Thomas G.
 APPLICANT: Sammons, R. Douglas
 APPLICANT: Shukla, Hridayabhinanjan
 APPLICANT: Wu, Kunsheng
 APPLICANT: Xu, Nanfei
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
 TITLE OF INVENTION: PLANTS
 FILE REFERENCE: 38-21(51936)A
 CURRENT APPLICATION NUMBER: US/60/207,458
 CURRENT FILING DATE: 2000-05-30
 NUMBER OF SEQ ID NOS: 152403
 SEQ ID NO 84816

```

; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: LIB3355-005-P1-K1-G2

```

US-60-207-458-84816

Query Match	6.0%	Score 66.6	DB 64	Length 469
Best Local Similarity	50.1%	Pred. No. 0.031		
Matches 193, Conservative	0	Mismatches 189	IndeIs 3	Gaps 1

[illegible]

RESULT 9
US-09-333-534-15000

```

sequence 15000, APPLICATION US/09335534
GENERAL INFORMATION:
APPLICANT: Conner, Timothy W.
APPLICANT: Ruan, Yijun G.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(15404)B
CURRENT APPLICATION NUMBER: US/09/333,534
CURRENT FILING DATE: 1999-06-08
NUMBER OF SEQ. ID NOS: 18718
SEQ ID NO 15000
LENGTH: 450

```

; ORGANISM: *Arabdopsis thaliana*
 ; OTHER INFORMATION: Clone ID: LIB25-082-Q1-E1-E12
 US-09-333-534-15000

```
Query Match          5.9%; Score 65; DB 17; length 450;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1037	CCCAATCTTAAAGCAAAAGCATTAGAAAAGGAGTGAGAGAGAGAAGAAAACTA	1096
Dd	55	CCCAATTTTAAAGCAAAGCCATTAGAAAAGGAAGTAGTGAGAGACAGAGAAACTA	114
QY	1097	GCTTC	1101
Dd	115	GCTTC	119

RESULT 10
US-09-654-617-121875
Sequence 121875, Application US/09654617
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Liu, Jingdong

RESULT 15
US-09-654-617-340019/c
; Sequence 340019, Application US/09654617
; GENERAL INFORMATION:

Search completed: June 11, 2003, 05:00:37
Job time : 2648.14 secs

100

Db	2342	CAGAATTGGCACAACAATGAAAGGAAATGTAATAATATTAACCTTTAATAATAATAAAT	2401
Oy	301	TGAGTAAATGTGTTTTCTGACTATTGAGGGCCAAAAAAGACAAATGCCAAAGTCTACG	360
Db	2402	TGAGTAAATGTGTTTTCTGACTATTGAGGGCCAAAAAAGACAAATGCCAAAGTCTACG	2461
Oy	361	GGTTTGACGTCCAGTGGGTGAATATATCTGTTTGGACGACGGTCGTG	420
Db	2462	GGTTTGACGTCCAGTGGGTGAATATATCTGTTTGGACGACGGTCGTG	2521
Oy	421	AGGGGTCCTTGACATTTTCAGTGTCTACCCCTACTCGTAGGCCACCCCTTTTCCAT	480
Db	2522	AGGGGTCCTTGACATTTTCAGTGTCTACCCCTACTCGTAGGCCACCCCTTTTCCAT	2581
Oy	481	ATCTGAAGGGTAATTTGGAAATCCCAATTAAACCGATTPAGACCCGACGGACTGCT	540
Db	2582	ATCTGAAGGGTAATTTGGAAATCCCAATTAAACCGATTPAGACCCGACGGACTGCT	2641
Oy	541	GGGATTCGTGAGACATTTATCAAAAATTATAGACGAATGGGTATATTAATTAAAA	600
Db	2642	GGGATTCGTGAGACATTTATCAAAAATTATAGACGAATGGGTATATTAATTAAAA	2701
Oy	601	ACTACAACTTGATCAGATTAATAATTTCTATAACACTTTACGATGGATCTGACATGTA	660
Db	2702	ACTACAACTTGATCAGATTAATAATTTCTATAACACTTTACGATGGATCTGACATGTA	2761
Oy	661	TCTAATGACTTTTTTTTTTCTACACAGGTGGATGAAAGTATATAGACTATTAGCCAGAGA	720
Db	2762	TCTAATGACTTTTTTTTTTCTACACAGGTGGATGAAAGTATATAGACTATTAGCCAGAGA	2821
Oy	721	CAATTGATTATAGATATATTCATTATATTCATATGATATTTATATTAATAGCTGTAAAC	780
Db	2822	CAATTGATTATAGATATATTCATTATATTCATATGATATTTATATTAATAGCTGTAAAC	2881
Oy	781	TATTTGACATCGAGCTTCTGCAACTTTGTTTTTAATTAGAGTTTAATAATAA	840
Db	2882	TATTTGACATCGAGCTTCTGCAACTTTGTTTTTAATTAGAGTTTAATAATAA	2941
Oy	841	AGTATTTAAAGAGACATATACGAGCCAAACAAAGTATAGACACGAGAAACAAAGCCAT	900
Db	2942	AGTATTTAAAGAGACATATACGAGCCAAACAAAGTATAGACACGAGAAACAAAGCCAT	3001
Oy	901	GAACCTCATGTTAGTTTAAGCTTAATAGACATTTATTAATTTTAATGACATGA	960
Db	3002	GAACCTCATGTTAGTTTAAGCTTAATAGACATTTATTAATTTTAATGACATGA	3061
Oy	961	TAAACAATATATTTCTGACTCTTTTAAACCCCTCTTACAAACAGAAAGTCCCTTTT	1020
Db	3062	TAAACAATATATTTCTGACTCTTTTAAACCCCTCTTACAAACAGAAAGTCCCTTTT	3121
Oy	1021	CAGTAGAAGTCCGATTCCTCAATCTTAAAGACAAAGCCATTGAAAGAAAGAGTAGAG	1080
Db	3122	CAGTAGAAGTCCGATTCCTCAATCTTAAAGACAAAGCCATTGAAAGAAAGAGTAGAG	3181
Oy	1081	AGAGAGAGAACTAGCTCC 1101	
Db	3182	AGAGAGAGAACTAGCTCC 3202	

```

RESULT 2
US-09-502-426A-1
: Sequence 1, Application US/09502426A
: GENERAL INFORMATION:
: APPLICANT: AZPIROZ, Ricardo
: APPLICANT: CHOE, Sungbwa
: APPLICANT: FELDMANN, Kenneth
: TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
: FILE REFERENCE: 2225-0001
: CURRENT APPLICATION NUMBER: US/09/502,426A
: CURRENT FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: 60/119,657
: PRIOR FILING DATE: 1999-02-11
: PRIOR APPLICATION NUMBER: 60/119,658

```

```

? PRIOR FILING DATE: 1999-02-11
? NUMBER OF SEQ ID NOS: 25
? SOFTWARE: FastSeq for Windows Version 4.0.0
? SEQ ID NO: 1
? LENGTH: 6888
? TYPE: DNA
? ORGANISM: Arabidopsis sp.
? US-09-502-426A-1

```

Query Match	100.0%;	Score 1101;	DB 7;	Length 6888;
Best Local Similarity	100.0%;	Pred. No. 8.9e+197;		
Matches 1101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1	TATGTACAAATATTAATATTAGTCATTAACATGCTATGAAAGTCCAAAAAATTT	60
Db	2102	AATGTACAAATATTAATATTAGTCATTAACATGCTATGAAAGTCCAAAAAATTT	2161
OY	61	TGTTTAAAGAAATCTCCAAATTTTTTTTTTATGGAACAAGAAATAACAGATGAAAC	120
Db	2162	TGTTTAAAGAAATCTCCAAATTTTTTTTTTATGGAACAAGAAATAACAGATGAAAC	2221
OY	121	TATTTTGTGTGGAAATGGAAATGATTAATATTCATTATAGCAAAATTTTAAAAATATATTA	180
Db	2222	TATTTTGTGTGGAAATGGAAATGATTAATATTCATTATAGCAAAATTTTAAAAATATATTA	2281
OY	181	GCCTATACGCGCTCAAAAGTATGTTATCTAGTAGGTGTAATTAATTAATCATGTGCGATT	240
Db	2282	GCCTATACGCGCTCAAAAGTATGTTATCTAGTAGGTGTAATTAATTAATCATGTGCGATT	2344
OY	241	CAGAAATGGGCAACAAATGAAACGGAAATTTAAATATTAATCTTTTAAATTAATAAAT	300
Db	2342	CAGAAATGGGCAACAAATGAAACGGAAATTTAAATATTAATCTTTTAAATTAATAAAT	2401
OY	301	TGAATTAATATGTTTTCTGACTATTTGAGGGGCAAAAAAACAATGCCAAAGCTACG	360
Db	2402	TGAATTAATATGTTTTCTGACTATTTGAGGGGCAAAAAAACAATGCCAAAGCTACG	2461
OY	361	GGTTTGACTGTCCAGTTCCGTAATATCTAATTAACCTGTCTTTGACCGACGCTGTGT	420
Db	2462	GGTTTGACTGTCCAGTTCCGTAATATCTAATTAACCTGTCTTTGACCGACGCTGTGT	2521
OY	421	AGGGGTCCCTTGTGACATTTTCACTGTTCTACCCCTACGCGAGAGCCACCCCTTTCCAT	480
Db	2522	AGGGGTCCCTTGTGACATTTTCACTGTTCTACCCCTACGCGAGAGCCACCCCTTTCCAT	2581
OY	481	ATCCTAAGGGTAATTTTGGAAATCCCAATTTTAAACCGATGAGACCGTACGGACTGCT	540
Db	2582	ATCCTAAGGGTAATTTTGGAAATCCCAATTTTAAACCGATGAGACCGTACGGACTGCT	2641
OY	541	GGGATTTCTGCGGAGCACTTATTCAAAAATTTTGTGACAGATGGGTTTATTAATTAA	600
Db	2642	GGGATTTCTGCGGAGCACTTATTCAAAAATTTTGTGACAGATGGGTTTATTAATTAA	2701
OY	601	ACTACAACTTGATCAGATTAATAAATTTTCAATAACACTTTTACGATGGATCTGACATCTA	660
Db	2702	ACTACAACTTGATCAGATTAATAAATTTTCAATAACACTTTTACGATGGATCTGACATCTA	2761
OY	661	TCTTATGACTTTTTTTTTTCTACACAGGTGGATGAAGTTTATGTACTATTAAGCAGAGA	720
Db	2762	TCTTATGACTTTTTTTTTTCTACACAGGTGGATGAAGTTTATGTACTATTAAGCAGAGA	2821
OY	721	CAATTTGATTAAGATATATCATTCATTAATCAGATATTTATATATTAATTAATCTGTTAAC	780
Db	2822	CAATTTGATTAAGATATATCATTCATTAATCAGATATTTATATATTAATTAATCTGTTAAC	2881
OY	781	TATTTTACGATCGCAGCTTCTGCAACTTTTGTTTTTAATTTAAGAGTTTAAATAATAA	840
Db	2882	TATTTTACGATCGCAGCTTCTGCAACTTTTGTTTTTAATTTAAGAGTTTAAATAATAA	2941
OY	841	AGTATTTAAAGAGACATATACGAGGCAACAAAGTAATGACACGAGAAACAAAAGCAT	900
Db	2942	AGTATTTAAAGAGACATATACGAGGCAACAAAGTAATGACACGAGAAACAAAAGCAT	3001

QY	901	GAAGTCATCTGGTAGTTAAAGCTTAATAAGACATTTTAAATTTTAATGACAGTGA	960
Db	3002	GAGCTCATCTGGTAGTTAAAGCTTAATAAGACATTTTAAATTTTAATGACAGTGA	3061
QY	961	TAACAAATTAATATTTTCTGACTTCTTTAAAGCCCGCTTACAAGACAGAGCTCCCTTTT	1022
Db	3062	TAACAAATTAATATTTTCTGACTTCTTTAAAGCCCGCTTACAAGACAGAGCTCCCTTTT	3123
QY	1021	CAGTAGAAGTCCGATTCCTCAATCTTTAAAGCAAAAGCCATTGAAGAAAGAGAAAGTAGTAG	1080
Db	3122	CAGTAGAAGTCCGATTCCTCAATCTTTAAAGCAAAAGCCATTGAAGAAAGAGAAAGTAGTAG	3181
QY	1081	AGAGAGAGAGAACTAGCTCC	1101
Db	3182	AGAGAGAGAGAACTAGCTCC	3202

```

RESULT 3
US-60-466-412-85418
: Sequence 85418, Application US/60466412
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: IAKOUBOVA, Olga
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CU001466
: CURRENT APPLICATION NUMBER: US/60/466,412
: NUMBER OF SEQ ID NOS: 429241
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 85418
: LENGTH: 16885
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-60-466-412-85418

```

[illegible]

RESULT 4
US-09-615-606A-35136/c
; Sequence 35136, Application US/09615606A
; GENERAL INFORMATION:

```

1  APPLICANT: Abbad, Mark S.
2  APPLICANT: Buehler, Robert E.
3  APPLICANT: Byrum, Joseph R.
4  APPLICANT: Coombs, Brian E.
5  APPLICANT: Heck, Gregory R.
6  APPLICANT: La Rosa, Thomas J.
7  APPLICANT: Nelson, Donald E.
8  APPLICANT: Shukla, Hridayabhiranjan
9  APPLICANT: Thompson, Michael D.
10 TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
11 TITLE OF INVENTION: Plants
12 FILE REFERENCE: 38-21(15444)C
13 CURRENT APPLICATION NUMBER: US/09/615,606A
14 CURRENT FILING DATE: 2000-07-13
15 NUMBER OF SEQ ID NOS: 91663
16 SEQ ID NO 35136
17 LENGTH: 458
18 TYPE: DNA
19 ORGANISM: glycine max
20 FEATURE:
21 NAME/KEY: unsure
22 LOCATION: (1)..(458)
23 OTHER INFORMATION: unsure at all n locations
24 FEATURE:
25 OTHER INFORMATION: Clone ID: LIB3053-011-Q1-N1-C9
26 US-09-615-606A-35136

```

Query Match 5.0%; Score 55; DB 6; Length 458;
Best Local Similarity 46.1%; Pred. NO. 0.41;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

[illegible]

```

RESULT 5
US-09-837-604A-1465/c
; Sequence 1465, Application US/09837604A
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: De la Pena, Robert C.
; APPLICANT: Bougri, Olegs
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51892)B
; CURRENT APPLICATION NUMBER: US/09/837.604A

```



```
LOCATION: (111)..(111)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (139)..(139)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (191)..(191)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (199)..(199)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (212)..(213)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (222)..(222)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (241)..(241)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (261)..(261)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (297)..(297)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (364)..(364)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (372)..(372)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (394)..(394)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc-feature
LOCATION: (418)..(418)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (424)..(424)
OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-293-46414

Query Match
Best Local Similarity 4.9%; Score 54.2; DB 6; Length 427;
Matches 137; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 28 ATAAACATGATAGAAAGTCCAAAAAATTTGTTAACAGAACTCCAAATTTT 87
DB 397 ATAAACATGATAGAAAGTCCAAAAAATTTGTTAACAGAACTCCAAATTTT 87
QY 88 TTTTATGAGAACAGAAATTAACAGATGAGAAACATTTTGTGGAATGGAAGTAA 147
DB 337 ATAAACATGATAGAAAGTCCAAAAAATTTGTTAACAGAACTCCAAATTTT 147
QY 148 TATACATTAAGCAATTTTAAAAATTTATATAGCCCTATACGGCTCAAGTGTATC 207
DB 277 TATATATTAAGCAATTTTAAAAATTTATATAGCCCTATACGGCTCAAGTGTATC 207
QY 208 TAGTAGGTGTAATTAATATGATGATGGTGCATTCACAAATTTGGACAAACATGAAAAACGA 267
```

```
DB 217 AATTNNVAAAAAATATATNTATTAANNTCAATTCATTAATAATATATTAATAAATA 158
QY 268 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 315
DB 157 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 110

RESULT 8
US-09-949-016-87162
Sequence 87162, Application US/09949016
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 87162
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-87162

Query Match
Best Local Similarity 4.9%; Score 54.2; DB 7; Length 601;
Matches 157; Conservative 1; Mismatches 154; Indels 1; Gaps 1;

QY 6 ACAATTTTAAATTTTATGTCATATACATGATGAGAAAGTCCAAAAAATTTGTTA 65
DB 212 ATAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 271
QY 66 ACAGAACTTCCAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 124
DB 272 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 331
QY 125 TTGTTGGAATGAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 184
DB 332 TTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 391
QY 185 ATACGGCTCAAGTATGTTATCTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 244
DB 392 ACATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 451
QY 245 ATGGGACAAATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 304
DB 452 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 511
QY 305 TAAATGCTTTTC 317
DB 512 AAGTGTCTTTTC 524

RESULT 9
US-10-311-455-776/c
Sequence 776, Application US/10311455
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
```


[illegible]

268 ATTAAATATTACCTTTAAATAA

12345678910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210310410510610710810911011111211311411511611711811912012112212312412512612712812913013113213313413513613713813914014114214314414514614714814915015115215315415515615715815916016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221321421521621721821922022122222322422522622722822923023123223323423523623723823924024124224324424524624724824925025125225325425525625725825926026126226326426526626726826927027127227327427527627727827928028128228328428528628728828929029129229329429529629729829930030130230330430530630730830931031131231331431531631731831932032132232332432532632732832933033133233333433533633733833934034134234334434534634734834935035135235335435535635735835936036136236336436536636736836937037137237337437537637737837938038138238338438538638738838939039139239339439539639739839940040140240340440540640740840941041141241341441541641741841942042142242342442542642742842943043143243343443543643743843944044144244344444544644744844945045145245345445545645745845946046146246346446546646746846947047147247347447547647747847948048148248348448548648748848949049149249349449549649749849950050150250350450550650750850951051151251351451551651751851952052152252352452552652752852953053153253353453553653753853954054154254354454554654754854955055155255355455555655755855956056156256356456556656756856957057157257357457557657757857958058158258358458558658758858959059159259359459559659759859960060160260360460560660760860961061161261361461561661761861962062162262362462562662762862963063163263363463563663763863964064164264364464564664764864965065165265365465565665765865966066166266366466566666766866967067167267367467567667767867968068168268368468568668768868969069169269369469569669769869970070170270370470570670770870971071171271371471571671771871972072172272372472572672772872973073173273373473573673773873974074174274374474574674774874975075175275375475575675775875976076176276376476576676776876977077177277377477577677777877978078178278378478578678778878979079179279379479579679779879980080180280380480580680780880981081181281381481581681781881982082182282382482582682782882983083183283383483583683783883984084184284384484584684784884985085185285385485585685785885986086186286386486586686786886987087187287387487587687787887988088188288388488588688788888989089189289389489589689789889990090190290390490590690790890991091191291391491591691791891992092192292392492592692792892993093193293393493593693793893994094194294394494594694794894995095195295395495595695795895996096196296396496596696796896997097197297397497597697797897998098198298398498598698798898999099199299399499599699799899910001001100210031004100510061007100810091010101110121013101410151016101710181019102010211022102310241025102610271028102910301031103210331034103510361037103810391040104110421043104410451046104710481049105010511052105310541055105610571058105910601061106210631064106510661067106810691070107110721073107410751076107710781079108010811082108310841085108610871088108910901091109210931094109510961097109810991100110111021103110411051106110711081109111011111112111311141115111611171118111911201121112211231124112511261127112811291130113111321133113411351136113711381139114011411142114311441145114611471148114911501151115211531154115511561157115811591160116111621163116411651166116711681169117011711172117311741175117611771178117911801181118211831184118511861187118811891190119111921193119411951196119711981199120012011202120312041205120612071208120912101211121212131214121512161217121812191220122112221223122412251226122712281229123012311232123312341235123612371238123912401241124212431244124512461247124812491250125112521253125412551256125712581259126012611262126312641265126612671268126912701271127212731274127512761277127812791280128112821283128412851286128712881289129012911292129312941295129612971298129913

268 ATTAAATATTACCTTAATAATAATTAATAATTGAGTA 306
.....CCATTTCTTAAAAAAATATATACCTT 6628

Db 6627 ATACATACATATACACGTACATAAATATATCTTTTA 6589

RESULT 14

```

US-09-947-916-332/c
: Sequence 332, Application US/09947916
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig
: TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON CHROMOSOMES
: TITLE OF INVENTION: 19, METHODS OF DETECTION, AND USES THEREOF
: FILE REFERENCE: C1001296
: CURRENT APPLICATION NUMBER: US/09/947,916
: CURRENT FILING DATE: 2003-03-20
: NUMBER OF SEQ ID NOS: 337
: SEQ ID NO 332
: LENGTH: 3297126
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1) ... (3297126)
: OTHER INFORMATION: n = A,T,C or G
US-09-947-916-332

```

RESULT 15
US-10-240

Sequence 131:Application US/10240453
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BEHLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
TITLE OF INVENTION: With DNA Transcription
FILE REFERENCE: 5013.1009
CURRENT APPLICATION NUMBER: US/10/240,453
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8

; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 350
 ; SEQ ID NO: 121

Search completed: June 11, 2003, 07:56:23
Job time : 822.985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 1669.1 Seconds

(Without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcaacaattataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	148	13.4	148	17	BH811044
c 2	145.6	13.2	473	17	AL768954 Arabidops
c 3	69.6	6.3	1101	17	CNS0039G
c 4	69.4	6.3	1201	17	CNS0167M
c 5	66.8	6.1	928	17	CNS00DKY
c 6	66.4	6.0	1201	17	CNS0165X

7	65.8	6.0	1101	17	CNS0021J	AL061936 Drosophila
8	65.2	5.9	1101	17	CNS00172	AL078714 Drosophila
c 9	64.6	5.9	1101	17	CNS00EVL	AL069706 Drosophila
c 10	63.6	5.8	1146	17	CNS021G2	AL176843 Tetradon
c 11	63.4	5.8	1101	17	CNS0021J	AL061936 Drosophila
c 12	63.2	5.7	1101	17	CNS00B01	AL057419 Drosophila
c 13	62.8	5.7	1009	17	CNS00587	AL057153 Drosophila
c 14	62.6	5.7	1101	17	CNS00EVL	AL057153 Drosophila
c 15	62.4	5.7	855	17	CNS04P5D	AL300874 Tetradon
c 16	62	5.6	1029	17	CNS01ZGM	AL174271 Tetradon
c 17	61.8	5.6	1101	17	CNS003BD	AL064091 Drosophila
c 18	61.6	5.6	609	17	CNS025K2	AL182171 Tetradon
c 19	61.6	5.6	1101	17	CNS0161L	AL106896 Drosophila
c 20	61.6	5.6	1190	17	CNS020MT	AL206908 Tetradon
c 21	61.6	5.6	1200	17	CNS016CO	AL106578 Drosophila
c 22	61.4	5.6	1092	17	CNS020K7	AL142826 Tetradon
c 23	61	5.5	910	17	CNS01G8P	AL142826 Tetradon
c 24	61	5.5	1085	17	AG075009	AG075009 Pan trogl
c 25	60.8	5.5	1101	17	CNS00B01	AL057419 Drosophila
c 26	60.8	5.5	1101	17	CNS001FB	AL060732 Drosophila
c 27	60.6	5.5	878	17	CNS0187R	AL108993 Drosophila
c 28	60.6	5.5	1101	17	CNS0039G	AL063921 Drosophila
c 29	60.2	5.5	1201	17	CNS016C2	AL106556 Drosophila
c 30	60	5.4	945	17	CNS04D0K	AL285149 Tetradon
c 31	59.8	5.4	1225	17	CNS0161D	AL106171 Tetradon
c 32	59.8	5.4	964	17	CNS07E8R	AL441457 T7 end of
c 33	59.8	5.4	966	17	CNS0052C	AL061951 Drosophila
c 34	59.6	5.4	1092	17	CNS020K7	AL175696 Tetradon
c 35	59.6	5.4	1368	10	BE420618	BE420618 HMM000.G0
c 36	59.2	5.4	524	17	CNS01090	AL167541 Tetradon
c 37	59.2	5.4	843	17	CNS00CS1	AL059666 Drosophila
c 38	59	5.4	987	17	CNS004PO	AL104456 Drosophila
c 39	59	5.4	994	17	CNS04NOJ	AL298972 Tetradon
c 40	58.8	5.3	1101	17	CNS008X3	AL052544 Drosophila
c 41	58.6	5.3	1001	17	CNS0155H	AL105023 Drosophila
c 42	58.6	5.3	1101	17	CNS00E07	AL069440 Drosophila
c 43	58.4	5.3	676	17	CNS003XU	AL065304 Drosophila
c 44	58.4	5.3	843	17	CNS00CS1	AL059666 Drosophila
c 45	58.4	5.3	996	17	CNS00F0H	AL071063 Drosophila

ALIGNMENTS

RESULT 1
LOCUS BH811044 148 bp DNA linear GSS 02-MAY-2002
DEFINITION SALK_057128 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_057128, DNA sequence.
BH811044
ACCSSION BH811044.1 GI:20388862
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.

thale cress,
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J., and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome
Unpublished (2001)

JOURNAL
COMMENT

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (Signal)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: eckeresalk.edu

This is single pass sequence recovered from the left border of
TDNA.

Class: T-DNA tagged.
Location/Qualifiers
1. 148
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="SALK_057128"
/clone.lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more T-DNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 46 a 28 c 14 g 60 t
ORIGIN

Query Match 13.4% Score 148; DB 17; Length 148;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ATGCACAGAAATACAGATAGAACTATTTGTTGTGAATGAGTAAATATAC 152
DB 148 ATGGAACAGAAATACAGATAGAACTATTTGTTGTGAATGAGTAAATATAC 89
QY 153 ATTAGCAATTTTAAAAATATATAGCCTATAGCGCTCAAGTATCTAGTA 212
DB 88 ATTAGCAATTTTAAAAATATATAGCCTATAGCGCTCAAGTATCTAGTA 29
QY 213 GGTGTATATATATGATGCTGCGATT 240
DB 28 GGTGTATATATATGATGCTGCGATT 1

RESULT 2 473 bp DNA linear GSS 18-JUN-2002
AL768954 Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
LOCUS genomic survey sequence.
ACCESSION AL768954
VERSION AL768954.1 GI:21522073
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
Strizhov, N., Li, Y., Rosso, M., Vlehoever, P., Dekker, K., Saedler, H. and Weisshaar, B.
TITLE A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines
JOURNAL Unpublished
AUTHORS Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics
JOURNAL Unpublished
AUTHORS Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone T3A5. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: <http://www.mpi-z-koeln.mpg.de/GABI-Kat/>.
FEATURES Location/Qualifiers
1. 473

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="GK-082A08-011867"
/clone.lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 13.2% Score 145.6; DB 17; Length 473;
Best Local Similarity 87.0%; Pred. No. 1.3e-17;
Matches 160; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 AATCTACAAATTTATATATTTAGCATATACATGCAATGCAAGTCCAAAAATTT 60
DB 290 AATCTACAAATTTATATATTTAGCATATACATGCAATGCAAGTCCAAAAATTT 349
QY 61 TGTACAGAAATTCCTCAATTTTTTTTTATGCAAGAAAGAAATACAGATAGAAAC 120
DB 350 TGTACAGAAATTCCTCAATTTTTTTTTATGCGCCCAAAATACCGATGGAACC 409
QY 121 TATTTGTGTGGAATGAGTAAATATATCATTAAGCAATTTTAAAAATTTATATA 180
DB 410 TTTTGGGGGGGAGAGGAGTAAATATATCTTAAACCAATTTTAAAAATTTTAA 469
QY 181 GCGT 184
DB 470 CCGT 473

RESULT 3 1101 bp DNA linear GSS 03-JUN-1999
CNS0039G Drosophila melanogaster genome survey sequence Tetr3 end of BAC #
LOCUS BACR08K10 of RPCR-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mammeter in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

```
/organism="Drosophila melanogaster"  
/db_xref="taxon:7227"
```

BASE COUNT ORIGIN	201 a	64 c	131 g	202 t	503 others

Query Match	6.38;	Score 69.6;	DB 17;	Length 1101;
Best Local Similarity	19.28;	Pred. No. 0.0016;		
Matches 97;	Conservative 218;	Mismatches 187;		

[illegible]

location/qualifiers
1. .1201

<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros.BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

BASE COUNT	323 a	87 c	79 g	551 t	161 others
ORIGIN					

Query Match	6.38;	Score 69.4;	DB 17;	Length 1201;
Best Local Similarity	37.28;	Pred. No. 0.0017;		
Matches 200; Conservative	57;	Mismatches 280;	Indels 0;	Gaps 0;

[illegible]

ORGANISM

KEYWORDS GSS.

ORGANISM

REFERENCE

REFERENCES

7777 E

THE
TODAY

JOURNAL

COMMENT

Db	1010	AATAAATAAAAAATTTT	TTTTTTTTTTTWT	TTTTTTTAANRAAWTTTWAMWAA	1060
RESULT 11	CNS0021J/C	CNS0021J	1101 bp	DNA	linear GSS 03-JUN-1999
LOCUS	DEFINITION	Drosophila melanogaster genome survey sequence TE73 end of BAC # BACR05H11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	VERSION	AL061936	1	GI:4940214	GSS.
SOURCE	ORGANISM	Drosophila melanogaster. Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	AUTHORS	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - web : www.genoscope.cns.fr)			
JOURNAL	TITLE	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the 1Sogenic strain Y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.			
FEATURES	SOURCE	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR05N11" /clone_1ld="RPCI-98" /note="end : TE73"			
BASE COUNT	ORIGIN	631 a	7 c	28 g	289 t 146 others
Query Match	Best Local Similarity	5.88;	Score 63.4;	DB 17;	Length 1101;
Matches	86; Conservative	25;	Mismatches	66;	Indels 0; Gaps 0;
QY	2	ATCTACAATTATTAATTTAGCATCAATGATGAGGAAGTCACAAAATAATTTT	61		
Db	742	WTTTAAATTAATTAATTAATTTTAAAMAAAAAATTAATAAAATTTTAAATTTW	683		
QY	62	GTTAACGAACACTCCAAATTTTGTATFARGAACAAGAATAACAGATGAAAAC	121		
Db	682	ATWATTAATAAAATTTATTTATTTTATTTATTAATTAATAAAATAAAATTTTAA	623		
QY	122	ATTATGGTGGAATGAGTAGTAATATACATTAAACCAATTTTAAAAAATTAATA	180		
Db	622	TTTTTTTWTAAATTAATAAAATTAATAAAATTTTATTAATAATTTTAAAAATTA	564		
RESULT 12	CNS00BO1/C	CNS00BO1	1101 bp	DNA	linear GSS 04-JUN-1999
LOCUS	DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR24D09 of RCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	VERSION	AL057419	1	GI:4937885	

KEYWORDS
SOURCE GSS.
ORGANISM Drosophila melanogaster.

REFERENCE
AUTHORS Drosophila melanogaster.
TITLE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
JOURNAL Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyridioidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.

COMMENT
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser. In Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCR-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR24D09"
/clone_1lb="RPCI-98"
/note="end : 77"

BASE COUNT
420 a 56 c 48 g 261 t 316 others

ORIGIN

Query Match Best Local Similarity Matches 176;	5.7%; Score 63.2; DB 17; Length 1101; 41.1%; Pred. NO.0.025; Conservative 51; Mismatches 199; Indels 2; Gaps 2;
--	---

```

558 TTATCAAAATTATTAGACGAATGGGTTTATTATAATTTAAAAAACCACAATCGATCAG 617
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
1048 TTTWTAETWTWATWATTTAATATWATTTTTAAAAAAAATWAAAWTTTATWAAAT 989
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
618 ATAAATTTCAATAAACACTTTAGACGAGATTCTGCAGACTATACATTTGCTTTTTT 677
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
988 TTAAATWTTTWWAATTTTWWATTTTWWATTTTAAATTTAATTTATTTTAAATTTWAAT 929
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
678 TTCACACGCGTGATGAAAGTAGTTAGTACTATTAGCCAGACAGAACAATGTATTAGATAT 737
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
928 WATFAAAAAAATAVA-AAWATTTTATTAATATTTWTTAATAATAATTTTWTATTATTWT 870
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
738 ATCATTAATTCATGATATTTATGATATAATATAGCTGTTAAACATTTACATCGCAGC 797
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
869 WTTATATATTWATTAATTTATATWAMWATTAATWTTTATTTTTTTTTTTTATTTWT 810
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
798 TTTCGCA-ACTTTGTTTAAATTTAAGTATTAATTAATAAAGATTTAAAGAGCA 856
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
809 TTTTMAATWATTTTWTTTTWTTTTMMATTAATTTAATTAATTAATTAAMWAAAAAAA 750
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
857 TAACGAGCAACAAGTAAATGAACAGCAGGAAAAAAGCATGAAGCTCATTTGGTTAG 916
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
749 AAAAAAAAAAAMWATAAAAAAATTTTTTTTTTATAASVMTVMSSSCAATVTWAMA 690
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
917 TTTTACGTTAATAAGATTTTATTAATTTTATAGCATGCTAATCAATTAATTTTC 976
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
Db 689 AATTAATSTMAATSVCASSBSVTTTATATTTTASAAMAMAKMAAATTTTSTGSTRT 630
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
OY 977 TGACTTCT 984
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
Db 629 TTAATTTT 622
      |||::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::

```

RESULT 13
CNS00587/c
LOCUS
DEFINITION
CNS00587 1009 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR11N01 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL057153
VERSION
AL057153.1 GI:4932541
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1.1009
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR11N01"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 263 a 116 c 96 g 477 t 57 others
ORIGIN
Query Match 5.7%; Score 62.8; DB 17; Length 1009;
Best Local Similarity 46.5%; Pred. No. 0.031;
Matches 148; Conservative 10; Mismatches 160; Indels 0; Gaps 0;
QY 1 AATCTCAAAATTTTAAATTTAGTCATACATGCAATGAAAGTTCCAAAAAATTT 60
DB 1006 AATATATAAANATWAAANATTAATTAATAAANATTAATAAANATTTTAAAT 947
QY 61 TGTAAACAGAACTTCCAAATTTTTTTTTTTATGACAGCAAGAAATTAACGATGA 120
DB 946 NATAAAAAANAAAAAANAAAAAATTAATTAATTAATTAATTAATTAATTAAT 887
QY 121 TATTTTGTGAGTAGAGTAGATTAATATACATTAAGCAAAATTTTAAAAATTA 180
DB 886 AATTTAAATATWAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 827
QY 181 GCCTATACGGCGCTCAAGATGTATCTAGTAGGTGAATTAATTAATGATGCGCAT 240
DB 826 TAAAAAANAAAAAANAAAAAANAAAAAATTAATTAATTAATTAATTAATTAAT 767
QY 241 CGAATTTGGAGACACATGAAGCAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 766 AATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 707
QY 301 TGAGTAATGTGTTTCT 318
DB 706 TTAATTAATTAATTTTWT 689

RESULT 14
CNS00591
LOCUS
DEFINITION
CNS00591 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL069706
VERSION
AL069706.1 GI:4949849
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1.1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN
Query Match 5.7%; Score 62.6; DB 17; Length 1101;
Best Local Similarity 35.1%; Pred. No. 0.032;
Matches 140; Conservative 65; Mismatches 194; Indels 0; Gaps 0;
QY 8 AATTTATTAATTTTATGTCATACATGCAATGCAAAATTTTGTAAAC 67
DB 587 AATWTATTAATTAATTAATTAATTAATTAATTTTAAATTAATTAATTAAT 646
QY 68 AGAACTTCCAAATTTTTTTTTTTATGACAGCAAGAAATTAACATGCAATTTTG 127
DB 647 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 706
QY 128 TTGAGGAGGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 187
DB 707 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 766
QY 188 CGCGCTCAAGATGTATGATGATGATGATGATGATGATGATGATGATGATGAT 247
DB 767 WTTWTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 826
QY 248 GGAACACAAATGAAGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 307
DB 827 WATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 886
QY 308 ATGTGTTTCTGCTATGAGGGGCAAAAAAGCAATCCCAAAAGTCTACGGGTTGA 367

[illegible]

```

QY      246 TTGGGCAACAAATGAAAAGGAATTAAAAATTAACCTTAAAAATAAATAATTTGAGT 305
          ||| | | | | | | | | | : | | | | | | | | | |
Db      711 ATATATAAAAAAAAAAAAAAAAAATAAATAATTAATTAATAAATAATTAATTTAAAA
          ||| | | | | | | | | | | | | | | | | | | | | |
QY      306 AAATGTGTTTTCTGACTATTGAGGGCAAAAAAAAAAGACAATGCCAAA 353
          ||| : | | | | | | | | | | | | | | | | | | | |
Db      771 AAAAMATATTTTTTTTNNATTTTATATAAAAAAAATAATTAATTAATTA 818
          ||| | | | | | | | | | | | | | | | | | | | | |

Search completed: June 11, 2003, 06:58:35
Job time : 1673.1 secs

```

```
Search completed: June 11, 2003, 06:58:35
Job time : 1673.1 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 953.95 Seconds
(without alignments)

10921.753 Million cell updates/sec

Title: US-09-502-426B-1_COPY_6111_6468

Perfect score: 358

Sequence: 1 aaaaaaaaaagatgaagf.....tgcacaatacacaacaaag 358

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	4838	8 AF044216	AF044216 Arabidops
2	356.4	99.6	84196	8 AF7345	AF132979 Arabidops
3	121	33.8	1691	8 AF412114	AF412114 Arabidops
4	58	16.2	5449	6 AX346543	AX346543 Sequence
5	58	16.2	7560	6 AX346125	AX346125 Sequence
6	57.2	16.0	173786	2 AC107582	AC107582 Rattus no
7	56.4	15.8	309233	2 AC098557	AC098557 Rattus no
8	56.4	15.8	5926	6 AX346554	AX346554 Sequence
9	55.6	15.5	12356	6 AX251264	AX251264 Sequence
10	55.6	15.5	169546	2 AC004157	AC004157 Plasmodu
11	55	15.4	7351	6 AX344930	AX344930 Sequence
12	55	15.4	20486	6 AX281500	AX281500 Sequence
13	54.8	15.3	93491	2 AC116967	AC116967 Dictyoste
14	54.8	15.3	349980	6 AX344570	AX344570 Sequence
15	54.6	15.3	123280	2 AC117076	AC117076 Dictyoste
16	54.6	15.3	268147	2 AC116966	AC116966 Dictyoste
17	54.2	15.1	2814	6 AX347029	AX347029 Sequence
18	54.2	15.1	6657	3 AC114263	AC114263 Dictyoste
19	54.2	15.1	67919	2 AC123610	AC123610 Mus muscu
20	54.2	15.1	204652	2 PFMA113P6	PFMA113P6 Plasmodu
21	54	15.1	7138	6 AX323841	AX323841 Sequence
22	53.2	14.9	187013	2 AC116920	AC116920 Dictyoste
23	53.2	14.9	6047	2 AC115582	AC115582 Dictyoste
24	53.2	14.9	14147	6 AX251501	AX251501 Sequence
25	53.2	14.9	14147	6 AX347392	AX347392 Sequence
26	53.2	14.9	14147	6 AX349113	AX349113 Sequence
27	53	14.8	7261	6 AX251424	AX251424 Sequence
28	53	14.8	133501	2 AC116956	AC116956 Dictyoste
29	53	14.8	349980	6 AX344567	AX344567 Sequence
30	52.8	14.7	5678	6 AX346040	AX346040 Sequence
31	52.8	14.7	107739	2 AC116979	AC116979 Dictyoste
32	52.8	14.7	256172	2 AC005139	AC005139 Plasmodu
33	52.8	14.7	310779	2 AC005140	AC005140 Plasmodu
34	52.6	14.7	6657	3 AC114263	AC114263 Dictyoste
35	52.6	14.7	24091	2 AC014610	AC014610 Dictyoste
36	52.6	14.7	106993	3 AE002751	AE002751 Drosophila
37	52.6	14.7	349980	6 AX344564	AX344564 Sequence
38	52.4	14.6	150903	2 AC113218	AC113218 Rattus no
39	52.2	14.6	6418	6 AX251768	AX251768 Sequence
40	52.2	14.6	6418	6 AX345224	AX345224 Sequence
41	52.2	14.6	16750	6 AX251068	AX251068 Sequence
42	52.2	14.6	16750	6 AX345423	AX345423 Sequence
43	52.2	14.6	66441	3 PFMA11P4	PFMA11P4 Plasmodu
44	52.2	14.6	234112	3 PFMA14P2	PFMA14P2 Plasmodu
45	51.8	14.5	627	11 CNS06EWH	AL395655 T7 end of

ALIGNMENTS

RESULT 1

LOCUS AF044216

DEFINITION Arabidopsis thaliana sterold 22-alpha-hydroxylase (DWF4) gene,

ACCESSION AF044216

VERSION AF044216.1 GI:2935341

KEYWORDS complete cds.

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 4818)

CHOE,S., DILKES,B.P., FUJIOKA,S., TAKATSUTO,S., SAKURAI,A. and

TITLE Feldmann, K. A.
 The DWF4 gene of Arabidopsis encodes a cytochrome P450 that
 mediates multiple 22alpha-hydroxylation steps in brassinosteroid
 biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE 2 (bases 1 to 4818)
 AUTHORS Choe, S., Dilkes, B. P., Azpilroz, R. and Feldmann, K. A.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona,
 Tucson, AZ 85721, USA
 FEATURES
 source Location/Qualifiers
 1. 4818
 /organism="Arabidopsis thaliana"
 /strain="Ws-2"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="8 cM from nga6"
 990. 4398
 /gene="DWF4"
 /note="dwarf4"
 /allele="dwarf4-1; dwarf4-2; dwarf4-3; dwarf4-4"
 990. 1055
 /gene="DWF4"
 /join(1133. 1353,1434. 1758,1844. 1996,2095. 2409,
 2562. 2654,2746. 2824,2931. 3040,3795. 4398)
 /gene="DWF4"
 /product="steroid 22-alpha-hydroxylase"
 /join(1133. 1353,1434. 1758,1844. 1996,2095. 2409,
 2562. 2654,2746. 2824,2931. 3040,3795. 4040)
 /gene="DWF4"
 /function="brassinosteroid biosynthetic enzyme"
 /note="member of the cytochrome P450 superfamily; CYP90B1"
 /codon_start=1
 /product="steroid 22-alpha-hydroxylase"
 /protein_id="AAC05093.1"
 /db_xref="GI:2935342"
 /translation="MEETEHTLLPLLLPSLLSLFLILKRRNRKTRNDLPKGS
 GMPFLDETIGYLPATYATLGDPMQOHVSXGKTYISNLEGEPTIVSADGLNRFTLO
 NEGRLEFESYPSRIGGILGKMSMLVGVDMHDMRSISLNFSLHARLRTLLDVERH
 TLEVLDSMOONSIFSAODEAKKFTFPMKHMNDGSEETOLKKEYTEFKGVAS
 APLNPGTAYHKLQSRATLTKFERKMERKDKIEDDEEVEVTEDEAKSKSH
 VKKQRTDDDLGLVNLKHSNSTQIIDLILSLPAGHETSVAIALAIFELQACPAV
 EELREHLEIARAKKELGSELNMDYKKMDFTQCVINETLRIGNVVRLHRRALDY
 RYKGYDIPSGMKVLPYISAVHLDNSRYDQNLFPNPMQONNGASSSGSSTGMN
 NYMPFGGPRLCAGSELAKLEMAVFTIHHLVILKFWELAEDDQPFARPEVDFPNGLPPIR
 VSRIL"
 BASE COUNT 1566 a 727 c 888 g 1637 t
 ORIGIN
 Query Match 100.0%; Score 358; DB 8; Length 4818;
 Best Local Similarity 100.0%; Pred. No. 5e-51;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGATGAAAGTATTTTATTTCTCTCTTTTGTGATAATTTAAATCA 60
 DB 4041 AAAAAAAAAAGATGAAAGTATTTTATTTCTCTCTTTTGTGATAATTTAAATCA 4100
 QY TTTTGTGGCCATGATATATAAATTGATAAATATATATATGATATCGTTT 120
 DB 4101 TTTTGTGGCCATGATATATAAATTGATAAATATATATATGATATCGTTT 4160
 QY TTAGTTCGGGTTGAGAAAAAGGTTTCGACTTTCGAAAGTGAGATGATATAGATTGG 180
 DB 4161 TTAGTTCGGGTTGAGAAAAAGGTTTCGACTTTCGAAAGTGAGATGATATAGATTGG 4220
 QY 181 GAGCTGAGTTGAGCTTTTGACATTTTGATATGATGATGTTGATATTAAGTGCACACT 240
 DB 4221 GAGCTGAGTTGAGCTTTTGACATTTTGATATGATGATGTTGATATTAAGTGCACACT 4280
 QY 241 ATTAACCTTAATGGGCTTCTATTAAGGCCCAATATATAGATTATACAAAGTGAC 300
 DB 4281 ATTAACCTTAATGGGCTTCTATTAAGGCCCAATATATAGATTATACAAAGTGAC 4340
 QY 301 AACTTTACTTCGTTTTTGATCCGACCAATATACAAATTTGCAATACCAACACAG 358
 DB 4341 AACTTTACTTCGTTTTTGATCCGACCAATATACAAATTTGCAATACCAACACAG 4398
 RESULT 2
 ATT3A5
 LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
 ACCESSION AL132979
 VERSION AL132979.2 GI:6782244
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 84196)
 Bloembergen, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetler, F. and
 Salanoubat, M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 84196)
 AUTHORS EU Arabidopsis sequencing project.
 JOURNAL Direct Submission
 TITLE Submitted (25-JUN-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
 bloem@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project
 Coordinator: Marcel Salanoubat and Francis Quetler, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 On Jan 27, 2000 this sequence version replaced gi:6434247.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 FEATURES
 source Location/Qualifiers
 1. 84196
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 2963. 3338
 /note="355 bp LTR"
 3339. 37564
 /note="LTR-transposon"
 7565. 79922
 /note="367bp LTR"
 12339. 13197
 /gene="T3A5.10"
 /join(12339. 12665,12742. 12803,12888. 13055,13125. 13197)
 /gene="T3A5.10"
 /note="similarity to cyclin-dependent kinase inhibitor
 (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB62432.1"
 /db_xref="GI:6561966"
 /translation="MAVRRRRRDYVENCVTITTVKRRKMEERVDLVESRIILSPCV
 QATNRGIVARNAGASEISVIVYRRDSPVEQCIIEEDSSVSCSSEKSKR
 IERVDLEENGGDDRETETSNITDPLNKSSESMNDSSVAVEDVESRRRLKSLHETV
 KEALEDLDFQVAKDRNLKLLKESMKNFPERDEPLGGGRYEMVKNLP"
 12339. 12665
 /gene="T3A5.10"
 /number=1
 12666. 12741
 /gene="T3A5.10"
 /number=1
 12742. 12803
 /gene="T3A5.10"
 /number=2
 12804. 12887
 Intron
 exon
 Intron
 exon
 Intron


```

|||||
Db 26032 TTTTGGCCCAATGATATATAAAATTGGATTAATATATATGATTCGTTT 25973
|||
Qy 121 TTAGTCGGGTTGAGAAAAGGTTTCAGCTTCGAAAGTGCAGATGATATAGATTGG 180
|||
Db 25972 TTAGTCGGGTTGAGAAAAGGTTTCAGCTTCGAAAGTGCAGATGATATAGATTGG 25913
|||
Qy 181 GACCTAGGTTGAGCTTTTGACATTTGTATGATGCTGTTGATTTAGTGCACACT 240
|||
Db 25912 GACCTAGGTTGAGCTTTTGACATTTGTATGATGCTGTTGATTTAGTGCACACT 25853
|||
Qy 241 ATTAACCTTAATGGCTTTCTATAGGCCCAATTATATAGATTATACAAAGTGAC 300
|||
Db 25852 ATTAACCTTAATGGCTTTCTATAGGCCCAATTATATAGATTATACAAAGTGAC 25793
|||
Qy 301 AACTTTACTCTGTTTTCGATCCGAAGCATATACAAATTCGCAATACCAACGACG 358
|||
Db 25792 AACTTTACTCTGTTTTCGATCCGAAGCATATACAAATTCGCAATACCAACGACG 25735
|||

RESULT 3
AF412114 1691 bp mRNA linear PLN 24-SEP-2001
LOCUS Arabidopsis thaliana AT3g50660/T3A5_40 mRNA, complete cds.
DEFINITION AF412114
VERSION AF412114.1 GI:15724347
KEYWORDS F1,CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 1691)
REFERENCE
AUTHORS Cheuk,R., Chen,H., Kim,C.-J., Koesema,E., Meyers,M.C., Banh,J.,
Bowser,L., Carlincl,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Arabidopsis cDNA clones
Unpublished
2 (bases 1 to 1691)
Cheuk,R., Chen,H., Kim,C.-J., Koesema,E., Meyers,M.C., Banh,J.,
Bowser,L., Carlincl,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (21-AUG-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
COMMENT
RIKEN Genomic Sciences Center (GSC) members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J., Bowser,L.,
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G.,
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,
Davis,R.W., Theologis,A., and Ecker,J.R.
Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to

```

```

this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.
FEATURES
source
1. 1691
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAFL06-76-F19(R1351)"
/note="ecotype: Columbia"
1. 28
5'/UTR
29..1570
/note="steroid 22-alpha-hydroxylase (DWF4)"
/codon_start=1
/product="AT3g50660/T3A5_40"
/protein_id="AAL06567.1"
/db_xref="GI:15724348"
/translation="MFEETHLLPLLLPSLILFLILKRNKRTRPNLPKGS
GMPPLGETITLKYPTTITGDRQGVSKRTKYSNLGEPTIYADAGLNFTILQ
NEGRLFECTSPKSTIGILGRWSMLVLGDHMRDRSISLNLSHARLRTILKVERH
TLFVLDSMOONSIFSAQDEAKKFTFNLMARHMSMDGEETDQKREYVFMGVS
APLNPETAYHAKLOSRAITLKEFERMEERKIDKEDEDEEVEKTEDEAKMSDH
VRKORTDDDLGWYLKHSNSTEOILDLILSLFAGHETSVAALAIFFLOACPKAY
EELREELLETARAKKEGESELMNDVKKMDPQCIVNLTIRIGNAVRFPHRKLKXY
RYGYDLPSEMKVLPVYSAYHLDNSRDQENLFRPMWQONNGASSGSGSESTWKN
NMFEGGPRLCAGESELAKLEMAVFHHLVLFKEWELAEDEKPPAFVDFPNCPLPR
VSRIL"
1571..1691
3'UTR
BASE COUNT 524 a 286 c 363 g 518 t
ORIGIN
Query Match 33.8% Score 121; DB 8; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2,7e-11;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAAGATGATATATAAAATTTGGATTAATATATGATTCGTTT 120
Db 1571 AAAAAAAAAAAGATGATATATAAAATTTGGATTAATATATGATTCGTTT 1630
Qy 61 TTTTGGCCCAATGATATATAAAATTTGGATTAATATATGATTCGTTT 120
Db 1631 TTTTGGCCCAATGATATATAAAATTTGGATTAATATATGATTCGTTT 1690
Qy 121 T 121
Db 1691 T 1691
RESULT 4
AX346543 5449 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 1614 from Patent WO0200928.
DEFINITION AX346543
ACCESSION AX346543.1 GI:18494429
VERSION
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Olek,A., Piepenbrock,C. and Beiln,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1614 03-JAN-2002;
Epidemiol. Infect.
FEATURES
source
1. 5449
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1146 a 90 c 1299 g 2914 t
ORIGIN
Query Match 16.2% Score 58; DB 6; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.68;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

QY 4 AAAAAAAAAAGATGATATTTTATTCCTCTTTTGGATATTTAAATCATTT 63
 DB 5256 AATTAATAATTAAGATATTTTGAATTTTATTTTATTTATTAATGATT 5315
 QY 64 TTTTCCCAATGATATTAATAAATTTGATTAATTAATTAATGATATTCGTTTGA 123
 DB 5316 TGGTTAAAGTTTATTAATATTTTATTAATATTTTATTAATTTGATATTTTATTT 5375
 QY 124 GTTCGGGTTT 133
 DB 5376 TTTTGGGTTT 5385

RESULT 5
AX346125

LOCUS AX346125 7560 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 1196 from Patent WO0200928.
 ACCESSION AX346125
 VERSION AX346125.1 GI:18494011
 KEYWORDS

SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1
 Olek, A., Piepenbrock, C. and Berlin, K.
 Diagnosis of diseases associated with the immune system
 Patent: WO 0200928-A 1196 03-JAN-2002;
 Epigenomics AG (DE)

FEATURES
 source
 Location/Qualifiers
 1..7560

/organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2154 a 47 c 1534 g 3805 t
 ORIGIN

Query Match 16.2%; Score 58; DB 6; Length 7560;
 Best Local Similarity 51.6%; Pred. No. 0.63;
 Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

QY 2 AAAAAAAAAAGATGATATTTTATTCCTCTTTTGGATATTTAAATCATTT 61
 DB 2728 AATTAATAATTAAGATATTTTGAATTTTATTTTATTTATTAATGATT 2787
 QY 62 TTTTCCCAATGATATTAATAAATTTGATTAATTAATTAATGATATTCGTTTGA 121
 DB 2788 TGGTTAAAGTTTATTAATATTTTATTAATATTTTATTAATTTGATATTTTATTT 2847
 QY 122 TACTTGGGTTTGAGAAAGGTTTCGACTTCGAAAGTGACGATATAGATTCGG 181
 DB 2848 TGTATTGTATTTAATAAATTAATTTTATTTTATTTTATTTTAAATTAATAT 2907
 QY 182 AGCTAGGTGAGTCTTTGACATTTTGTATGATGT---TGTGATTATTAAGTTCGAC 237
 DB 2908 TATTTATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTT 2967
 QY 238 ACTATTAACCTTAATGAGCTTCTATTAAGCCCAATTAATTAATTAATTAATTA 297
 DB 2968 AATTTTGAATTAATTAATTTTATTTTATTTTAAAGAAATTAATTAATTAAGA 3027
 QY 298 GACACCTT 305
 DB 3028 GATTAATT 3035

KEYWORDS
SOURCE
ORGANISMREFERENCE
AUTHORS

HTG: HTGS, PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 173786)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alshrooks, S.L., Amaralunga, H.C., Aye, J.R., Ayale, M., Banks, T.,
 Barbara, J., Benton, J., Bimge, K., Blankenburg, K., Bonin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Checko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gatti, R.,
 Correll, J.H., Guevara, W., Gunatane, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homst, F., Howard, S., Huber, J., Huij, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C.,
 Katslovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, M., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinley, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, R., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, S., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nokenko, S., Ogun, K., Okunodu, G.,
 Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherren, S., Scott, G., Shen, H., Shoochthai, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Sytek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
 2 (bases 1 to 173786)
 Direct Submission
 Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 173786)
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18846355.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GNF

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

RESULT 6
 AC107582 173786 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-29B17, *** SEQUENCING IN PROGRESS
 DEFINITION *** 97 unordered pieces.
 ACCESSION AC107582
 VERSION AC107582.3 GI:21736513

Assembly program: Phrap; version 0.990329
Consensus quality: 85701 bases at least Q40
Consensus quality: 94901 bases at least Q30
Consensus quality: 103067 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 92 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1044: contig of 1044 bp in length
1045 1144: gap of unknown length
1145 2264: contig of 1120 bp in length
2265 2364: gap of unknown length
2365 3650: contig of 1286 bp in length
3651 3750: gap of unknown length
3751 4927: contig of 1177 bp in length
4928 5027: gap of unknown length
5028 6190: contig of 1163 bp in length
6191 6290: gap of unknown length
6291 7306: contig of 1016 bp in length
7307 7406: gap of unknown length
7407 8489: contig of 1083 bp in length
8490 8589: gap of unknown length
8590 9778: contig of 1189 bp in length
9779 9878: gap of unknown length
9879 11183: contig of 1305 bp in length
11184 11283: gap of unknown length
11284 12388: contig of 1105 bp in length
12389 12488: gap of unknown length
12489 13543: contig of 1055 bp in length
13544 13643: gap of unknown length
13644 14803: contig of 1160 bp in length
14804 14903: gap of unknown length
14904 16009: contig of 1106 bp in length
16010 16109: gap of unknown length
16110 17310: contig of 1201 bp in length
17311 17410: gap of unknown length
17411 18416: contig of 1006 bp in length
18417 18516: gap of unknown length
18517 19745: contig of 1229 bp in length
19746 19845: gap of unknown length
19846 20945: contig of 1100 bp in length
20946 21045: gap of unknown length
21046 22132: contig of 1087 bp in length
22133 22232: gap of unknown length
22233 23496: contig of 1264 bp in length
23497 23596: gap of unknown length
23597 24728: contig of 1132 bp in length
24729 24828: gap of unknown length
24829 25982: contig of 1154 bp in length
25983 26082: gap of unknown length
26083 27212: contig of 1130 bp in length
27213 27312: gap of unknown length
27314 28536: contig of 1224 bp in length
28537 28636: gap of unknown length
28637 29865: contig of 1229 bp in length
29866 29965: gap of unknown length
29966 31115: contig of 1150 bp in length
31116 31215: gap of unknown length
31216 32311: contig of 1096 bp in length
32312 32411: gap of unknown length
32412 33460: contig of 1049 bp in length
33461 33560: gap of unknown length
33561 34759: contig of 1199 bp in length
34760 34859: gap of unknown length
34860 35937: contig of 1078 bp in length
35938 36037: gap of unknown length

36038 37158: contig of 1121 bp in length
37159 37258: gap of unknown length
37259 38454: contig of 1196 bp in length
38455 38554: gap of unknown length
38555 39786: contig of 1232 bp in length
39787 39886: gap of unknown length
39887 40915: contig of 1029 bp in length
40916 42118: contig of 1103 bp in length
42119 42218: gap of unknown length
42219 43243: contig of 1025 bp in length
43244 43343: gap of unknown length
43344 44733: contig of 1390 bp in length
44734 44833: gap of unknown length
44834 46747: contig of 1914 bp in length
46748 46847: gap of unknown length
46848 48885: contig of 2038 bp in length
48886 48985: gap of unknown length
48986 50103: contig of 1118 bp in length
50104 50203: gap of unknown length
50204 51246: contig of 1043 bp in length
51247 51346: gap of unknown length
51347 52475: contig of 1129 bp in length
52476 52576: gap of unknown length
52576 54247: contig of 1671 bp in length
54247 55700: contig of 1354 bp in length
55701 55800: gap of unknown length
55801 57340: contig of 1540 bp in length
57341 57440: gap of unknown length
57441 58866: contig of 1426 bp in length
58867 58966: gap of unknown length
58967 60262: contig of 1296 bp in length
60263 60362: gap of unknown length
60363 62046: contig of 1684 bp in length
62047 62146: gap of unknown length
62147 63851: contig of 1705 bp in length
63852 63951: gap of unknown length
63952 65247: contig of 1296 bp in length
65248 65347: gap of unknown length
65348 66839: contig of 1492 bp in length
66840 66939: gap of unknown length
66940 68082: contig of 1143 bp in length
68083 68182: gap of unknown length
68183 69518: contig of 1336 bp in length
69519 69619: gap of unknown length
69620 71049: contig of 1431 bp in length
71050 71150: gap of unknown length
71150 72450: contig of 1301 bp in length

Query Match 16.0%; Score 57.2; DB 2; Length 173786;
Best Local Similarity 62.2%; Pred. No. 0.4;
Matches 89; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGAGTATTTTATCTCTTTTATTTTATTTGATTAATTTAAATCA 60
Db 27835 AAAAAAAAAAAAAAAAAATATATTTTATTTTATTTTATTTGAAAAAATAATTT 27894
QY 61 TTTTTCCTCCATGATATTAATAATTGATTAATAATATATGATATTCGTTT 120
Db 27895 ATTTTATTTTATTTTATTTTACAAAAAATAAAAAAAAAATTTTATTTTATTTTAT 27954
QY 121 TTATTCGGGTTTGAGAAAGG 143
Db 27955 TTATTAANGGGGGGAAAGG 27977

RESULT 7
AC098557 309233 bp DNA linear HTG 12-JUL-2002
LOCUS Rattus norvegicus clone CH230-81p10, *** SEQUENCING IN PROGRESS
DEFINITION *** 61 unordered pieces.
ACCESSION AC098557

```

VERSION      AC098557.6  GI:21728854
KEYWORDS     HTGS; HTGS_Phrasel.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1 (bases 1 to 309233)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaralung,H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbieri,J., Benton,J., Blimage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burke,C., Burrell,K.L., Byrd,N.C.,
             Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
             Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C., Hollins,B.,
             Homs,F., Howard,S., Huber,S., Hulik,S., Hume,J., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
             Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
             Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
             Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
             Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,
             Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
             Nguyen,N., Nickerson,E., Nwokenwu,S., Ogun,M., Okunnu,G.,
             Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peety,U., Perez,L.,
             Peters,L., Pickens,R., Primus,E., Pu,U.L., Quiles,M., Ren,Y.,
             Rives,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,
             Scherger,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
             Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H.,
             Sutton,A., Svatek,A., Tabors,P., Tamerisa,A., Tamerisa,K., Tang,H.,
             Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
             Umani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
             Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
             Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
             Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstein,G. and Gibbs,R.
TITLE        Unpublished
JOURNAL      Direct Submission
REFERENCE    2 (bases 1 to 309233)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (24-Oct-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 309233)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (12-Jul-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Jul 11, 2002 this sequence version replaced gi:20066030.
             ----- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             ----- Project Information
             Center project name: CH230
             Center clone name: CH230-81p10
             ----- Summary Statistics
             Sequencing vector: Plasmid;

```

Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 233229 bases at least Q40
 Consensus quality: 233927 bases at least Q30
 Consensus quality: 234451 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Gendank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 61 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1 1069: contig of 1069 bp in length
* 1169: gap of unknown length
* 1170
* 1170: contig of 1354 bp in length
* 2524
* 2623: gap of unknown length
* 2624
* 3832: contig of 1209 bp in length
* 3833
* 3932: gap of unknown length
* 4955: contig of 1023 bp in length
* 4956
* 5055: gap of unknown length
* 6248: contig of 1193 bp in length
* 6348: gap of unknown length
* 6349
* 7523: contig of 1175 bp in length
* 7524
* 7623: gap of unknown length
* 7624
* 8765: contig of 1142 bp in length
* 8766
* 8865: gap of unknown length
* 10089: contig of 1224 bp in length
* 10090
* 10189: gap of unknown length
* 10190
* 11339: contig of 1150 bp in length
* 11340
* 11440: gap of unknown length
* 12630: contig of 1191 bp in length
* 12631
* 12730: gap of unknown length
* 13871: contig of 1141 bp in length
* 13872
* 13971: gap of unknown length
* 13972
* 14978
* 15077: gap of unknown length
* 15078
* 16258: contig of 1181 bp in length
* 16259
* 16358: gap of unknown length
* 17421: contig of 1063 bp in length
* 17422
* 17521: gap of unknown length
* 17522
* 18539: contig of 1018 bp in length
* 18540
* 18639: gap of unknown length
* 18640
* 19856: contig of 1217 bp in length
* 19857
* 19956: gap of unknown length
* 20981: contig of 1025 bp in length
* 20982
* 21081: gap of unknown length
* 22354: contig of 1273 bp in length
* 22355
* 22454: gap of unknown length
* 23840: contig of 1386 bp in length
* 23841
* 23940: gap of unknown length
* 25442: contig of 1502 bp in length
* 25443
* 25442: gap of unknown length
* 25543
* 26794: contig of 1252 bp in length
* 26795
* 26895: gap of unknown length
* 28026: contig of 1132 bp in length
* 28027
* 28126: gap of unknown length
* 29239: contig of 1113 bp in length
* 29240
* 29339: gap of unknown length
* 30567: contig of 1228 bp in length
* 30568
* 30667: gap of unknown length
* 31810: contig of 1143 bp in length
* 31811
* 31910: gap of unknown length
* 31911
* 32522: contig of 1342 bp in length
* 33253
* 33352: gap of unknown length
* 33353
* 34543: contig of 1191 bp in length
* 34544
* 34643: gap of unknown length
* 35643: contig of 1000 bp in length
* 35644
* 35743: gap of unknown length
* 36743: contig of 1000 bp in length

```



```
* 36744 36843: gap of unknown length
* 36844 38114: contig of 1271 bp in length
* 38115 38214: gap of unknown length
* 38215 39335: contig of 1121 bp in length
* 39336 39435: gap of unknown length
* 39436 40648: contig of 1213 bp in length
* 40649 42073: gap of unknown length
* 42074 42173: gap of unknown length
* 42174 43290: contig of 1117 bp in length
* 43291 43390: gap of unknown length
* 43391 44446: contig of 1056 bp in length
* 44447 44545: gap of unknown length
* 44546 45742: contig of 1196 bp in length
* 45743 45842: gap of unknown length
* 45843 46961: contig of 1119 bp in length
* 46962 47061: gap of unknown length
* 47062 48199: contig of 1138 bp in length
* 48200 48299: gap of unknown length
* 48300 49421: contig of 1122 bp in length
* 49422 49521: gap of unknown length
* 49522 50769: contig of 1248 bp in length
* 50770 50869: gap of unknown length
* 50870 52645: contig of 1776 bp in length
* 52646 52745: gap of unknown length
* 52746 54068: contig of 1323 bp in length
* 54069 54168: gap of unknown length
* 54169 55823: contig of 1655 bp in length
* 55824 55923: gap of unknown length
* 55924 56971: contig of 1048 bp in length
* 56972 57071: gap of unknown length
* 57072 58564: contig of 1493 bp in length
* 58565 58664: gap of unknown length
* 58665 60294: contig of 1630 bp in length
* 60295 60394: gap of unknown length
* 60395 61440: contig of 1046 bp in length
* 61441 61540: gap of unknown length
* 61541 63038: contig of 1498 bp in length
* 63039 63138: gap of unknown length
* 63139 64816: contig of 1678 bp in length
* 64817 64916: gap of unknown length
* 64917 66388: contig of 1472 bp in length
* 66389 66488: gap of unknown length
* 66489 67608: contig of 1120 bp in length
* 67609 67708: gap of unknown length
* 67709 68749: contig of 1041 bp in length
* 68750 70135: gap of unknown length
* 70136 70235: contig of 1286 bp in length
* 70236 76486: gap of unknown length
* 76487 76586: contig of 6251 bp in length
```

Query Match 15.8%; Score 56.6; DB 2; Length 309233;
Best Local Similarity 63.2%; Pred. No. 0.44;
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGATGAAATGATTTTATCTCTCTTTTGTGATAATTTAAATCA 60
DB 45162 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 61 TTTTGTGCGCATGATATATAAATTTGATATAATATATATGATATTCGTTT 120
DB 45222 TTTTGTGCGCATGATATATAAATTTGATATAATATATATGATATTCGTTT 120
QY 121 TTAGTTCGGGTTGAG 136
DB 45282 TTTTGTGCGGTTGAG 45297
```

RESULT 8
AX346554
LOCUS AX346554
DEFINITION Sequence 1625 from Patent WO0200928.
ACCESSION AX346554
5926 bp DNA linear PAT 01-FEB-2002

```
VERSION AX346554.1 GI:18494440  
KEYWORDS synthetic construct.  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Olek.A., Piepenbrock.C. and Berlin.K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 1625 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES  
source location/Qualifiers  
1..5926  
/organism="synthetic construct"  
/db_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 1562 a 55 c 1318 g 2991 t  
ORIGIN
```

Query Match 15.8%; Score 56.4; DB 6; Length 5926;
Best Local Similarity 48.2%; Pred. NO. 1.2;
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

```
QY 4 AAAAAAAAAAGATGATTTTATCTCTCTTTTGTGATAATTTAAATCAATT 63
DB 3171 AAGATGAGGTCGAGTGAATGTTTATTTTATTTTGTGATATTTATAG 3230
QY 64 TTTTGGCCATGATATATAAATTTGATATAATTTGATATTCGTTTATA 123
DB 3231 TTTGAATATGATATTTATATAGTTTGTGATTTTGAATTTAAATATTTAT 3290
QY 124 GTTCGGGTTGAGAAAGGTTTGACCTTTCGAAAGTGACGATATATGATGGAG 183
DB 3291 TATGAGATATTAAGTTTATGTTTGAAGAAAGATTTTGAATGAAATTTTGGGCTG 3350
QY 184 CTAGTTCGAGCTTTCGACATTTGATTTGATTTGATTTAGTCGACACATTT 243
DB 3351 GGAATTTTATGATTTTATTTTATTTTGTTCGATTTGATTTATTTAGTTTAT 3410
QY 244 AATCCTTAATGAGCTTCTATTAAGCCCATTTATTTATGATTTATTAACAAGTGACAC 303
DB 3411 TTAACGAAGATTTGGGTTGTTGAGAGATGGCGTGGATGGAAGTATTTAGTAT 3470
QY 304 TTTTACCTTCGTTTGTGATCCGACATTA 333
DB 3471 TATATTTGATTTTGAATGATTTAATAA 3500
```

RESULT 9
AX251264
LOCUS AX251264
DEFINITION Sequence 232 from Patent WO0168912.
ACCESSION AX251264
VERSION AX251264.1 GI:15984687
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Olek.A., Piepenbrock.C. and Berlin.K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
JOURNAL Patent: WO 0168912-A 232 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
source location/Qualifiers
1..12356
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3645 a 68 c 2127 g 6516 t
ORIGIN

Query Match 15.5%; Score 55.6; DB 6; Length 12356;
Best Local Similarity 53.2%; Pred. No. 1.4;

COMMENT	FEATURES	BASE COUNT	ORIGIN
On Aug 12, 2000 this sequence version replaced gi:8810447.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 2 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			
* 1 23466: contig of 23466 bp in length			
* 23467 23666: gap of unknown length			
* 23667 169546: contig of 145880 bp in length.			
Location/Qualifiers			
1. 169546			
/organism="Plasmodium falciparum"			
/db_xref="taxon:5833"			
/chromosome="12"			
/clone="PPYAC293"			
/clone="3D7"			
69871 a 15381 c 15705 g 68389 t 200 others			

	Query Match	15.5%	Score 55.6	DB 2	Length 169546	
	Best Local Similarity	51.6%	Pred. No. 0.74			
	Matches	127	Conservative	0	Mismatches 119	Indels 0; Gaps 0;
OY	1	AAAAAAAAAAGATGAAGAATTTTTCCTCCTTTTTTTTGATAATTGTAATCA	60			
Dd	136495	AAAAAAAAAAAAAAGATPAAACATTTGAACCTTTTGTATGTTTACCCTTTTTTTTT	1365			
OY	61	TTTTTTTGGCCCATGATATATATAAAATTTGGATAAAATAATATATTTGCATATTCGTTTT	120			

QY 20 TATTATTATCTCTCTCTTTTTTTTTTTCGATAATTAAATCAATTTTTTTTGGCCAAAGATA 79
Db TTTGGTATTTGGTATTCGGTATGCTTTAGAAAGCTTTTTTAAATTTTGGTTTTTTCGACATA 3824
QY 80 TATTAATAATTTGGCATTAATAATATTATTGGATATATCGTTTTTATAGTCGGTTTGAGAAA 139
Db ATATTATATTATTAATTTTATAGAGTTAAATTTATATTTTGGTTTTTTTAAATTTTTTAA 3884
QY 140 AGGGTTTCACATTCGAAAGTGCAGACATATATATAGATTTGGAGCTAGTTGAGCTCTTTG 139
Db AGSTTTTAGGTTTGGGTATGTTGTTTTTATGTTTATGTTTATGAGTAAATTTGCAATTTTGT 3944
QY 200 GACATTTGTAATTCGACATGTTCTGCATTAATTAATTAAGTCGCACACTATTAACCTTAATGGGCT 259
Db AATAAATATGTTGGTATTTGTAATTTTAAATATAGTACTTTTAATATGTAATTTAAATTTGTA 4004
QY 260 TTCTATATAGGCCCAATTATATATAGCATTAATAACAAGTGACACATTTACTCTGTTTTTG 319
Db TTTATATATAGAGCTGTAATATTGTAATTTTAAATTAAGTACTGTCTTTTTTGTGTTTTGT 4064
QY 320 ATCCGAAGCAATTAACAATTTGCAAT 346
Db TTTAATATAGCATGTAAATTTTTTACT 4091

RESULT 12			
AX281500	AX281500	20486 bp	linear
LOCUS	Sequence	164 from Patent WO017376.	
DEFINITION	AX281500		
ACCESSION	AX281500.1	GI:16608755	
VERSION			

KEYWORDS synthetic construct.
SOURCE ORIGIN artificial sequences.

REFERENCE 1
Olek A., Piepenbrock C. and Berlin K.
Diagnosis of diseases associated with metastasis
Patent: WO 0177376-A 164 18-OCT-2001;
Epigenomics AG (DE)

FEATURES
source Location/Qualifiers
1..20486
/oranism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 6101 a 360 c 446 g 957 t 2 others

ORIGIN

Query Match 15.4%; Score 55; DB 6; Length 20486;
Best Local Similarity 49.6%; Pred. No. 1.6;
Matches 169; Conservative 0; Mismatches 170; Indels 2; Gaps 1;

OY 17 AAGTATTTTATTCCTCTCTTTTTTGTGTAATTAAATCAATTTTTGGCCCATG 76
| | | | | | | | | | | | | | | | | | | |
DB 11836 ATGATAGATGATTTAATAATTTAGAAATATGTTTTTAGTGTTTTTTTAAAA 11895
| | | | | | | | | | | | | | | | | | | |
OY 77 ATATATAAATTTTGGAATATATATTTGATATCGTTTGTGGGTGAC 136
| | | | | | | | | | | | | | | | | | | |
DB 11896 GTAAAGAAAATTTAGATATATATATATAGAATTTTGTATTTAAATGATTTT 11955
| | | | | | | | | | | | | | | | | | | |
OY 137 AAAAGGTTTCGATTCGAAAGTGAGCAGATGATATAGATGGAGCTGACT 196
| | | | | | | | | | | | | | | | | | | |
DB 11956 TTGTTTTTCGAATTTGGTAAAGTTTAATATATTTGATGTTTTTTTATTT 12015
| | | | | | | | | | | | | | | | | | | |
OY 197 TTGCACATTTGTATGATGTTGTTGATATTTAGTGCACTATTTAACTTAAATG 256
| | | | | | | | | | | | | | | | | | | |
DB 12016 TTACAGATTTAATAGTTATAGTTTGTAGATGATTAATAATAA--TAATTT 12073
| | | | | | | | | | | | | | | | | | | |
OY 257 GCTTCTATATAGGCCCAATTTATTTAGCATTTATACAAAGTGACACTTTTACTGTTT 316
| | | | | | | | | | | | | | | | | | | |
DB 12074 GATTTAAATAGATTTCGTAATATAGTATTTAAATTTAAATTTTAAATTTGTTT 12133
| | | | | | | | | | | | | | | | | | | |
OY 317 TTTGATCCGAACAATATGTCAAATACCAACACAA 357
| | | | | | | | | | | | | | | | | | | |
DB 12134 TTTATGTTMAAAATATTAATTTATTTTGAATTTAA 12174
| | | | | | | | | | | | | | | | | | | |

RESULT 13
AC116967/c AC116967 93491 bp DNA linear HTG 04-APR-2002
LOCUS Dictyostelium discoideum chromosome 2 map 5401525-5495014 strain AX4,
DEFINITION *** SEQUENCING IN PROGRESS ***. In ordered pieces.
ACCESSION AC116967
VERSION AC116967.1 GI:19920066
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
AUTHORS 1 (bases 1 to 93491)
Gloeckner,G., Eichinger,L., Szafrański,K., Pachebat,J., Dear,P.,
Leimann,R., Baumgart,C., Parra,G., April,J.F., Gilgo,R., Kumpf,K.,
Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 93491)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT CBS predictions from GenBank may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)

and the University Cologne, Institute for Biochemistry I
(<http://www.uni-koeln.de/dictyostelium/project.shml>)
Funding Agency : Deutsche Forschungsgemeinschaft (DFG).
NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source Location/Qualifiers
1..93491
/oranism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="5401525-5495014"

BASE COUNT 36436 a 9904 c 10019 g 37032 t 100 others

ORIGIN

Query Match 15.3%; Score 54.8; DB 2; Length 93491;
Best Local Similarity 50.4%; Pred. No. 1.2;
Matches 134; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAGATGGAAGATTTTATCTCTTTTTTTTGTATATTTTAAATCA 60
| | | | | | | | | | | | | | | | | | | |
DB 65656 AAAAAATAAAAACTCAATCAATCTTAACATTACATTTTCCTTTTAATTTCTTTT 65599
| | | | | | | | | | | | | | | | | | | |
OY 61 TTTTGTGCCCCATGATATATAAAAAATTTGATAAATAATATATGATATCGTTT 120
| | | | | | | | | | | | | | | | | | | |
DB 65596 TCTTCTATATCTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 65533
| | | | | | | | | | | | | | | | | | | |
OY 121 TTGTTGGGCTTTGCAAAAAGGTTTCGACTTTCGAAATGACAGATGATATACATTG 180
| | | | | | | | | | | | | | | | | | | |
DB 65536 TTGTGATCTAATATTTAATTTGATTTGACAGGCTAATATTTTATTTTATCTTTT 65477
| | | | | | | | | | | | | | | | | | | |
OY 181 GACCTGAGTTGACTCTTTGACATTTGATTTGATTTGATTTGATTTGATTTGATTTG 240
| | | | | | | | | | | | | | | | | | | |
DB 65476 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 65417
| | | | | | | | | | | | | | | | | | | |
OY 241 ATTAAACCTTAATGGCTTTCATA 266
| | | | | | | | | | | | | | | | | | | |
DB 65416 AATTAATATATTTATTTGCTTTTATTTTA 65391
| | | | | | | | | | | | | | | | | | | |

RESULT 14
AX344570 AX344570 349980 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 21 from Patent WO0200932.
DEFINITION AX344570
ACCESSION AX344570
VERSION AX344570.1 GI:18492456
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.

REFERENCE 1
Olek A., Piepenbrock C. and Berlin K.
Diagnosis of known genetic parameters within the mhc
Patent: WO 0200932-A 21 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source Location/Qualifiers
1..349980
/oranism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.548.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length of seq 2: 3.673778 <223>-split as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 21:59:33 ; Search time 81.9537 Seconds

(without alignments)
9837.455 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358

Sequence: 1 aaaaaaaaaagatgaagc.....tgcataaccacacacag 358

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2165239 seqs, 1125999159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	358	100.0	6888	21	AAA59599
2	58	16.2	5449	24	ABL33641
3	58	16.2	7560	24	ABL33223
4	56.4	15.8	5926	24	ABL33652
5	55.6	15.5	12356	22	AA545510
6	55	15.4	7351	24	ABL34028
7	55	15.4	20486	24	ABL34611
8	54.2	15.1	2814	24	ABL34127
9	54	15.1	7138	24	ABK28455

10	53.2	14.9	14147	22	AA546743	Tumour suppressor
11	53.2	14.9	14147	24	ABK33955	Human DNA for stag
12	53	14.8	7261	22	AA546670	Tumour suppressor
13	52.8	14.7	5678	24	ABL33138	Human immune syste
14	52.6	14.7	21231	23	ABL19002	Drosophila melanog
15	52.2	14.6	6418	24	AA561073	Human immune syste
16	52.2	14.6	6418	22	AA561073	Human immune regulat
17	52.2	14.6	16750	22	AA546314	Tumour suppressor
18	52.2	14.6	16750	24	ABL32521	Human immune syste
19	51.6	14.5	6145	24	ABL32972	Human immune syste
20	51.6	14.4	17131	24	ABL33053	Human immune syste
21	51	14.2	13584	24	ABL32615	Human immune syste
22	50.8	14.2	5311	24	ABL33019	Human immune syste
23	50.8	14.2	8087	22	ABL32742	Human immune syste
24	50.8	14.2	13606	22	AA545457	Chemically pretrea
25	50.8	14.2	13606	22	AA545561	Tumour suppressor
26	50.8	14.2	13606	24	ABL33810	Human immune syste
27	50.8	14.2	13606	24	ABL33810	DNA transcription
28	50.6	14.1	5649	22	AA546384	Tumour suppressor
29	50.6	14.1	5649	24	ABK40008	Human chemically p
30	50.6	14.1	5649	24	ABL32849	Human immune syste
31	50.4	14.1	5163	24	ABL33248	Human immune syste
32	50.2	14.0	14615	22	AA546704	Tumour suppressor
33	49.8	13.9	6134	22	AA546563	Human immune syste
34	49.8	13.9	6759	24	ABL32738	Human immune syste
35	49.8	13.9	8866	22	AA545433	Chemically pretrea
36	49.8	13.9	8866	24	ABK28280	DNA transcription
37	49.8	13.9	15649	24	ABL70544	Chemically treated
38	49.8	13.9	113515	24	ABL34175	Human immune syste
39	49.6	13.9	9347	24	ABL49336	Human polynucleoti
40	49.6	13.9	11422	24	ABK39936	Human chemically p
41	49.6	13.9	11422	24	ABL33218	Human immune syste
42	49.6	13.9	12639	24	ABN80106	Human immune syste
43	49.6	13.9	13427	24	ABL33926	Human chemically m
44	49.4	13.8	6609	24	ABL70526	Human immune syste
45	49.4	13.8	6609	24	AA561123	Chemically treated

ALIGNMENTS

RESULT 1	AAA59599	strand: DNA; 6888 BP.
ID	AAA59599	standard: DNA; 6888 BP.
AC	AAA59599;	
XX		
DT	14-NOV-2000	(first entry)
DE	DNA encoding a cytochrome P450 enzyme designated DMF4.	
XX		
KW	DMF4: cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	plant phenotype; cell elongation; ss.	
XX		
OS	Arabidopsis sp.	
XX		
PH	Key	Location/Qualifiers
FT	promoter	1..3203
FT		/*tag= a
FT	TATA_signal	3060..3125
FT		/*tag= b
FT	CDS	3203..6110
FT		/*tag= c
FT		/*product= "DMF4"
FT		/*note= "contains introns"
FT	exon	3203..3423
FT		/*tag= d
FT	intron	3424..3503
FT		/*tag= e
FT	exon	3504..3828
FT		/*tag= f
FT	intron	3829..3913
FT		/*tag= g

```

FT exon 3914..4066
FT /tag- h
FT Intron 4067..4164
FT /tag- 1
FT exon 4165..4479
FT /tag- 1
FT Intron 4480..4631
FT /tag- k
FT exon 4632..4724
FT /tag- 1
FT Intron 4725..4815
FT /tag- m
FT exon 4816..4894
FT /tag- n
FT Intron 4895..5000
FT /tag- o
FT exon 5001..5110
FT /tag- p
FT Intron 5111..5864
FT /tag- q
FT exon 5865..6110
FT /tag- r
FT 3'UTR 6011..6468
FT /tag- s

XX WO20047715-A2.
XX
XX 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-US03820.
XX
XX 11-FEB-1999; 99US-0119657.
XX 11-FEB-1999; 99US-0119658.
XX
XX (ARIZ-) ARIZONA BOARD OF REGENTS.
XX
XX PI Azpiroz R, Choe S, Feldmann KA;
XX DR WPI: 2000-549142/50.
XX DR P-PSDB; AAB07921.
XX
XX PT New isolated dwf4 polynucleotide useful for altering the phenotype of
XX plants, for diagnostic assays and in the production of antibodies -
XX
XX PS Claim 3; Fig 10A-G; 113pp; English.
XX
XX CC The present sequence encodes a DWF4 polypeptide. The polypeptide is a
XX cytochrome P450 enzyme that mediates multiple steps in synthesis of
XX brassinosteroids. Specifically, it mediates multiple
XX 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
XX polynucleotide is used for altering the phenotype of a plant. DWF4
XX plants display a dramatic reduction in the length of different organs,
XX and this size reduction is attributable to a defect in cell elongation.
XX The DWF4 polynucleotides and polypeptides can be used in diagnostic
XX assays and to generate antibodies, which can be used to produce
XX immunogenic compositions.
XX
XX SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
XX
XX Query Match 100.0%; Score 358; DB 21; Length 6888;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-55;
XX Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGATGAAGTATTTATCTCTCTTTTGTGATTAATTTAAATCA 60
XX |||||||
XX Db 6111 AAAAAAAAAAGATGAAGTATTTATCTCTCTTTTGTGATTAATTTAAATCA 6170
XX
XX QY 61 TTTTTCGCCCAATGATATATAAATTTGGATAAATATATTATTCGATTCGTTT 120
XX |||||||
XX Db 6171 TTTTTCGCCCAATGATATATAAATTTGGATAAATATATTATTCGATTCGTTT 6230
XX
XX QY 121 TTGTCGGGTTGAGAAAGGTTTCGATTCGAAAGTGACGATGATATAGATTGG 180
XX |||||||

```

```

Db 6231 TTAGTTCGGGTTTGAGAAAAGGTTTCGACTTTCGAAAGTGACGATATATAGATTGG 6290
QY 181 GAGCTAGGTTGAGCTTTTGACATTTGATGATGTTGTTGATTTATTTAGTTCGACACT 240
|||
Db 6291 GAGCTAGGTTGAGCTTTTGACATTTGATGATGTTGTTGATTTATTTAGTTCGACACT 6350
QY 241 ATTAAACCTTAAGTGGCTTTCTATTAAGCCCAATTATATATATATTAACAAAGTGAC 300
|||
Db 6351 ATTAAACCTTAAGTGGCTTTCTATTAAGCCCAATTATATATATATTAACAAAGTGAC 6410
QY 301 AACCTTACTGCTTTTGAGCCGAAAGCAATTAACAAATTCGAAATTAACAAAGTGAC 358
6411 AACCTTACTGCTTTTGAGCCGAAAGCAATTAACAAATTCGAAATTAACAAAGTGAC 6468

RESULT 2
ABL33641
ID ABL33641 standard; DNA; 5449 BP.
XX
XX ABL33641;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1614.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antihaemic; cytosolic; nocrotic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX OS Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX PA Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2002-130909/17.
XX
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX
XX PS Claim 1; SEQ ID NO 1614; 32pp + Sequence Listing; German.
XX
XX CC The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX SQ Sequence 5449 BP; 1146 A; 90 C; 1299 G; 2914 T; 0 other;
XX
XX Query Match 16.2%; Score 58; DB 24; Length 5449;
XX Best Local Similarity 65.4%; Pred. No. 0.053;
XX Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
XX
XX QY 4 AAAAAAAAAAGATGAAGTATTTATCTCTCTTTTGTGATTAATTTAAATCAATT 63

```


Db 5256 AAAAATAAAGTATTTTGAATTTTATTTTATTTTATTTTAACTT 5315
OY 64 TTTTGGCCCAATGATATATTAATAATTTGATTAATATTTGATTTCTTTTAA 123
Db 5316 TGTATAAGTTTATTAATATTTTATTAATATTTTAATTTGATTTTATTTT 5375
OY 124 GTTCGGCTTT 133
Db 5376 TTTTGGCTTT 5385

RESULT 3
ABL33223
ID ABL33223 standard; DNA: 7560 BP.
XX
AC ABL33223;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1196.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
PI Olek A. Piepenbrock C, Berlin K.
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 1196; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7560 BP; 2154 A; 47 C; 1554 G; 3805 T; 0 other;

Query Match 16.2%; Score 58; DB 24; Length 7560;
Best Local Similarity 51.6%; Pred. No. 0.053;
Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

OY 2 AAAAAAAAAAGTGAATTTTATCTCTCTTTTATTTTATTTTAAATCAT 61
Db 2728 ATAAATTCAGATAAAGTAATTTTATTTTATTTTATTTAGATTATTTA 2787

OY 62 TTTTGGCCCAATGATATATTAATAATTTGATTAATATTTGATTTCTTTT 121
Db 2788 TTTTATGTGAGATTTTAAATTTTAAAGATAATTTTATTTTATGCTGTAFTT 2847
OY 122 TAGTTCGGCTTTGAGAAAAGGCTTTCGACTTCCAAAGGACATGATATGATTGGC 181
Db 2848 TGTATTTGATTAATAAATTTATTTAGATTTTATTTGATTTTAAAAAATTAAT 2907
OY 182 AGCTAGGTGAGTCTTGGACATTTGATTTGATGTGATTTTATTTATTTTAAATTT 2967
Db 2908 TATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTT 2967
OY 238 ACTATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297
Db 2968 ATATTTCGATTTAAATTAATTTTATTTTAAAGAAATTAATTAATTAATTAAGA 3027
OY 298 GACAACTT 305
Db 3028 GATAAATT 3035

RESULT 4
ABL33652
ID ABL33652 standard; DNA: 5926 BP.
XX
AC ABL33652;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1625.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
PI Olek A. Piepenbrock C, Berlin K.
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 1625; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5926 BP; 1562 A; 55 C; 1318 G; 2991 T; 0 other;

Query Match 15.8%; Score 56.4; DB 24; Length 5926;
 Best Local Similarity 48.2%; Pred. No. 0.1;
 Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

4 AAAAAAGATGAAGATTTTATCTCTTTTGTGATTTTAAATCATTT 63
 3171 AAGATGAGGTCGATTTGTTTATTTTATTTTGAATTTTATGAG 3230
 64 TTTTGGCCATGATATATATAATTTGATTAATATTCGTTTAA 123
 3231 TTTGAATGTTATTTATAGATTTTGTGTTGGATTTTGAATTAATTTAT 3290
 124 GTTCGGGTTGAGAAAAGGCTTCGACTTCGAAAGTGACGATGATATGATGGAG 183
 3291 TATGAGATATTAAGTTTATGTTTGAAGAAGATTTTGAATGAAATTTGGGTG 3350
 184 CTAGGTTGACCTTTTGACATTTTGATTTGTTGATTTATGTCGACATTT 243
 3351 GGAGTTTATATATTTATTTATTTTGTGATTTGATTTGATTTTATTTTAAAT 3410
 244 AATCTTAATGAGGCTTTCTATAAGGCCAATTAATTAATTAACAAAGTACAC 303
 3411 TTAACGAAGTGGGTGGTGGAGATGGGTAGGAGGTTGGAAGATATCTAGTAT 3470
 304 TTTTACTCGTTTGTGATCCGAACATTA 333
 3471 TATATTTGATTTTGAATGATTTTAAATTA 3500

Db 3471 TATATTTGATTTTGAATGATTTTAAATTA 3500

RESULT 5
 AAS46510
 ID AAS46510 standard; DNA: 12356 BP.
 AAS46510:
 18-DEC-2001 (first entry)
 Tumour suppressor gene derived chemically modified sequence #232.
 Human; tumour suppressor gene; oncogene; antitumour; cytosolic;
 cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 cytosine methylation; ds.
 Homo sapiens.
 WO200168912-A2.
 20-SEP-2001.
 15-MAR-2001; 2001MO-EP02955.
 15-MAR-2000; 2000DE-1013847.
 06-APR-2000; 2000DE-1019058.
 07-APR-2000; 2000DE-1019173.
 30-JUN-2000; 2000DE-1032529.
 01-SEP-2000; 2000DE-1043826.
 (EPIG-) EPIGENOMICS AG.
 Olek A, Piepenbrock C, Berlin K;
 WPI; 2001-602752/68.
 Fragments of chemically modified genes associated with tumour suppressor
 genes and oncogenes, useful in designing primers and probes for
 analysing diseases associated with cytosine methylation state e.g.
 cancer
 Claim 1; SEQ ID NO 232; 27pp; English.
 The invention relates to a nucleic acid comprising a sequence of 18
 bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC complementary sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 12356 BP; 3645 A; 68 C; 2127 G; 6516 T; 0 other;
 CC
 CC Query Match 15.5%; Score 55.6; DB 22; Length 12356;
 CC Best Local Similarity 53.2%; Pred. No. 0.14;
 CC Matches 118; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

11 AGATGAAGATTTTATCTCTCTTTTGTGATTAATTAATCAATTTTTC 70
 9051 AGATTAATGCTAGTTATTTATTTTATTTTGAAGATTTTAAATGTTTATAGT 9110
 71 CCATGATATTAATTAATTTGATTAATTAATTAATGATTCGTTTATGTTGGG 130
 9111 GCTTGATGATTTATTTATTTTAAATTAATTAATGAGGTTTATTTTATTTATGT 9170
 131 TTTGAAAAGGCTTTCGACTTTGCAAGTGACAGTATATATGATTTGGAGCTAGGT 190
 9171 TATTAGATTTGTTATAGTTTCTTTTGGATATAGATTAATTAATTTGGTAGATGAT 9230
 191 GAGTCTTGGACATTTGATTTGATGATGTTGATTTAGTG 232
 9231 ATTTATTTGATTTGATTTGATTTTATTTTGGATGATTAATG 9272

Db 9231 ATTTATTTGATTTGATTTGATTTTATTTTGGATGATTAATG 9272

RESULT 6
 ABL32028
 ID ABL32028 standard; DNA: 7351 BP.
 ABL32028:
 26-MAR-2002 (first entry)
 Human immune system associated gene SEQ ID NO: 1.
 Human; immune system disease; cytosine methylation; antiasthmatic;
 antiatherosclerotic; antianemic; cytosolic; neutropenic;
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 antineumatic; antiallergic; antidiabetic; antiparasitic;
 antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 gene; ds.
 Homo sapiens.
 WO200200928-A2.
 03-JAN-2002.

XX	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX	gene; ds.
OS	Homo sapiens.
XX	
PN	MO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Plegenbrock C, Berlin K;
DR	WPI: 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 2100; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
XX	Sequence 2814 BP; 629 A; 144 C; 770 G; 1271 T; 0 other;

	CC	compared to another set of genetic and/or epigenetic parameters, the
	CC	differences serving as basis for diagnosis and/or prognosis events which
	CC	are disadvantageous to patients. The present sequence is one of the
	CC	533 genomic sequences derived from tumour suppressor genes and
	CC	oncogenes.
	CC	Note: The sequence data for this patent did not form part
	CC	of the printed specification, but was obtained in electronic
	CC	format directly from WIPO at
	CC	ftp.wipo.int/pub/published_pct_sequences.
	XX	
SQ	Sequence	14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
Query Match	14.9%;	Score 53.2; DB 22; Length 14147;
Best Local Similarity	48.18;	Pred. No. 0.36; Mismatches 163; Indels 0; Gaps 0;
Matches 151;	Conservative 0;	Mismatches 163; Indels 0; Gaps 0;
QY	7	AAAAAGATGAAAGCTATTTTATTCCTCTTTTTTTTGAGAAATTTTAAATCATTTTTT 66
Db	10811	AAAAGAATTGGTTAGTTGGTTTGTGTAATTTTGTTTTTTTTCGTGGTTTTTTTT 10870
QY	67	TTCGCCAAGATATAAAAAATTGGATAAATAATATATATGGAATTCGTTTTTACGT 126
Db	10871	TTTTTTTCGAATTAATAGTAGTATTTTTTTTTTTTATATTTGTAGGATTTTTTAGTT 10930
QY	127	CGGGTTTGAAAAGAGGTTTCGACTTCCGAAGTGCAGCAGCAATATATACATTTGGAGCTA 186
Db	10931	TTGGGTTTTTAAATACGTTGTCGTTTAAATTTTGTTTGTGTTTAAATATTTTACGAG 10990
QY	187	GATTGAGTCTTTTGACATTTTGTATTTGCAGTTCGTGATTAATAGTGCACACATTAANA 246
Db	10991	TTTATGTTTTTATTTTTTTTTTTGGTTTGTGTTAGTTGTTTTTTTTTGGCATTAATTTT 11050
QY	247	CCTTAAATGGGCTTCTAATAGGCCCAATTATATATAGATTATACAAAGTACAACCTT 306
Db	11051	TTTTTGATTTTTTATGTAAGTAAGTAATAGTTTTTTTTTGTGATTTTTTTTTTAATTT 11110
QY	307	TACTTCGTTTTTGA 320
Db	11111	TTAATTATTAATTTGA 11124
RESULT 11		
ABK33955		
ID	ABK33955	standard; DNA; 14147 BP.
XX	AC	
XX	ABK33955;	
DT	18-JUN-2002	(first entry)
XX		
DE	Human DNA for staging of Astrocytomas #20.	
XX		
KW	Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;	
KW	bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;	
KW	matrix assisted laser desorption/ionization mass spectrometry.	
XX		
OS	Homo sapiens.	
PN	WO200202808-A2.	
PD	10-JAN-2002.	
PF	02-JUL-2001; 2001WO-EP07538.	
PR	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
XX		
PA	(EPIC-) EPIGENOMICS AG.	
XX		
PI	Olek A, Piepenbrock C, Berlin K;	
DR	WFI; 2002-171649/22.	
PT	Novel chemically modified genomic DNA sequences, useful in the	

PT characterisation, classification, differentiation, grading, staging,
PT treatment and/or diagnosis of astrocytomas or predisposition to
PT astrocytomas
XX
PS Claim 1; SEQ ID No 39; 37pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence (I) of at
CC least 18 bases in length of a segment of chemically pre-treated genomic
CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
CC complement. Also included are an oligonucleotide or peptide nucleic
CC acid (or set thereof) of at least 9 nucleotides which hybridises to (I),
CC primers for (I), probes for detecting cytosine methylation or single-
CC nucleotide polymorphisms (SNP) in (I), an array of oligomers
CC or peptide nucleic acids for analysing diseases associated with the
CC methylation states of the CpG dinucleotides of (I). The array is useful
CC for determining genetic and/or epigenetic parameters, classification,
CC differentiation, grading, staging, treatment and/or diagnosis of
CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
CC methylations, involves obtaining a biological sample containing genomic
CC DNA, extracting the genomic DNA, converting cytosine bases which are
CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or
CC another base which is dissimilar to cytosine in terms of hybridisation
CC behaviour, by chemical treatment and amplifying chemically pre-treated
CC genomic DNA fragments using the array and a polymerase, where the
CC amplicates carry a detectable label. The method further involves
CC identifying methylation status of one or more cytosine positions, and
CC analysing methylation status of the cytosine positions by reference to
CC one or more data sets. The genomic DNA is chemically treated by using a
CC bisulphite, hydrogen sulphite or disulphite. The amplification
CC step amplifies DNA which is of particular interest in astrocytoma or
CC brain tissue, based on the specific genomic methylation status of brain
CC tissues, as opposed to background DNA. The amplicates carry a
CC fluorescent label or radionuclide. Optionally, the labels of the
CC amplicates are detachable molecule fragments having a typical mass
CC which are detected in a mass spectrometer. The fragments of chemically
CC pre-treated genomic DNA to be amplified, have a single positive or
CC negative charge for a better detectability in the mass spectrometer.
CC Preferably, the amplicates or fragments of the amplicates are
CC detected by matrix assisted laser desorption/ionization mass spectrometry
CC (MALDI) or using electron spray mass spectrometry (ESI). The
CC present sequence is one of the chemically pre-treated reference DNA
CC samples of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
Query Match 14.9%; Score 53.2; DB 24; Length 14147;
Best Local Similarity 48.1%; Pred. No. 0.36;
Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 7 AAAAGATGAAGATTTTATCTCTCTTTTTCATATTTTAAATCATTTT 66
DB 10811 AAAAGTTTGTTAGTTAGTTGTTTCTATTTTGTGTTTGTGTTT 10870
QY 67 TTCGCCAATGATATATAAAATTTGGATTAATATTAATGATTCGTTT 126
DB 10871 TTTTTCGAAATTAATGTTATTTTATTTTATGTTATGTTATTTTGGTT 10930
QY 127 CGGGTTTGAGAAAAGGTTTCGACTTCGAAAGTGACATGATATAGATGGACCTA 186
DB 10931 TTGGGTTTAAATGCGTTGCTGTTTATTTGGTTTGTGTTTAAATTTACGAG 10990
QY 187 GGTTCAGTCTTTGACATTTGATATGATGTTGATTTGTCGACACATATAA 246
DB 10991 TTTAGTTTATTTTATTTTGTGTTTGTGTTTGTGTTTGTGCGCATATTTT 11050
QY 247 CCTTAATAGGCTTCTATAGAGCCCAATTAATATAGATTAACAAAGTCAACTT 306
DB 11051 TTTTTCATTTTATGTAAGAAAGTAATAGTTTGTGTTTGTGTTTATATATTT 11110

QY 307 TACTTCGTTTGTGA 320
DB 11111 TTATATATATTTGA 11124
RESULT 12
AAS46670
ID AAS46670 standard; DNA; 7261 BP.
XX
AC AAS46670;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #392.
XX
KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A. Piepenbrock C, Berlin K;
XX
XX WPI: 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer
PS
PS Claim 1; SEQ ID No 392; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligonucleotide (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 7261 BP; 1862 A; 171 C; 1854 G; 3374 T; 0 other;

Query Match 14.8%; Score 53; DB 22; Length 7261;

Best Local Similarity 50.6%; Pred. No. 0.41; Mismatches 125; Indels 0; Gaps 0;

Matches 128; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 13 ATGAAGATATTTTATCTCTCTTTTGTATATATTTTAAATCATTTTGTGCCC 72
DB 6407 ATAAAGCAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 6466
OY 73 AATGATATATATATTTGATATATATATATATATATATATATATATATATAT 132
DB 6467 CGTAT 6526
OY 133 TGAGAAAGGCTTCGACCTTCGAAAGAGAGATATATATATATATATATATAT 192
DB 6527 TGTAGCAATATATATAGAGCTGCGATTTTGTAGTGTGATATATATATAT 6586
OY 193 GTCTTTGACATTTGTATTTGATGTTGATTTATTTAGTGTGACATATTTAA 252
DB 6587 TTTTATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 6646
OY 253 ATGGGCTTTCTAT 265
DB 6647 TTTGTTTATTTT 6659

RESULT 13

ABL33138
ID ABL33138 standard; DNA; 5678 BP.

AC ABL33138;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1111.

DE Human immune system associated gene; cytosine methylation; antiasthmatic;
XX Human system disease; antianemic; cytosine methylation; nocturnal;
XX antileukemic; anti-HIV; anticonvulsant; ophthalmological;
XX antineoplastic; antineoplastic; antidiabetic; arteriosclerosis; anaemia;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

XX Claim 1; SEQ ID NO 1111; 32bp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

Sequence 5678 BP; 1387 A; 32 C; 1284 G; 2975 T; 0 other;

Query Match 14.7%; Score 52.8; DB 24; Length 5678;

Best Local Similarity 54.0%; Pred. No. 0.44; Mismatches 92; Indels 0; Gaps 0;

Matches 108; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

OY 36 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 95
DB 3308 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3367
OY 96 AAT 155
DB 3368 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3427
OY 156 AAGTGCAGATGATATATATATATATATATATATATATATATATATATAT 215
DB 3428 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3487
OY 216 GTTGTGATTTATTTAGTGTG 235
DB 3488 TATGTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3507

RESULT 14

ABL19002
ID ABL19002 standard; DNA; 21231 BP.

AC ABL19002;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 8479.

XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO20011042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions

XX Claim 1; SEQ ID NO 8479; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (AB57737-AB872072).

XX The sequence data for this patent did not form part of the printed

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 17.0513 Seconds
(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaaagt.....tgtaaatatcaaacacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCITS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Dackfilles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	42.8	12.0	1493	1 US-08-340-820-24	Sequence 24, Appl
2	42.8	12.0	1493	1 US-08-593-535-24	Sequence 24, Appl
3	41.8	11.7	998	4 US-09-122-400B-5	Sequence 5, Appl
4	41	11.5	2230	4 US-08-378-313-24	Sequence 24, Appl
5	40.8	11.4	6243	2 US-09-056-075-1	Sequence 1, Appl
6	40.8	11.4	19557	5 PCR-US92-06300-1	Sequence 1, Appl
7	40.6	11.3	1733	3 US-09-073-569-1	Sequence 1, Appl
8	40.6	11.3	3095	6 5231168-1	Patent No. 5231168
9	40.4	11.3	10607	1 US-08-078-090-3	Sequence 3, Appl
10	40	11.2	2755	1 US-08-749-522-2	Sequence 2, Appl
11	39.4	11.0	289	1 US-08-341-568-3	Sequence 3, Appl
12	39.4	11.0	289	1 US-08-911-020-3	Sequence 3, Appl
13	39.4	11.0	19124	2 US-08-487-826B-13	Sequence 13, Appl
14	39	10.9	1895	4 US-09-444-336-7	Sequence 7, Appl
15	39	10.9	4285	4 US-09-410-464-1	Sequence 1, Appl
16	39	10.9	9048	3 US-08-973-273-4	Sequence 547, App
17	38.8	10.8	835	4 US-08-998-416-547	Sequence 498, Appl
18	38.8	10.8	863	4 US-08-998-416-498	Sequence 4, Appl
19	38.8	10.8	3138	1 US-07-867-106-4	Sequence 24, Appl
20	38.2	10.7	2836	3 US-08-747-221B-24	Sequence 26, Appl
21	38.2	10.7	2836	4 US-08-747-221B-26	Sequence 24, Appl
22	38.2	10.7	2836	4 US-09-005-051-24	Sequence 26, Appl
23	38.2	10.7	2836	4 US-09-005-051-26	Sequence 2, Appl
24	38.2	10.7	5852	1 US-07-867-106-2	Sequence 207, App
25	38.2	10.7	29604	3 US-08-781-891-207	Sequence 2606, Ap
26	38	10.6	552	4 US-09-134-001C-2606	Sequence 701, App
27	38	10.6	701	4 US-08-998-416-701	

28	38	10.6	2555	2 US-08-693-457-3	Sequence 3, Appl
29	38	10.6	2555	4 US-09-265-731-3	Sequence 3, Appl
30	38	10.6	8700	2 US-08-392-625-16	Sequence 16, Appl
31	38	10.6	8700	2 US-08-466-961A-16	Sequence 16, Appl
32	38	10.6	8700	2 US-08-645-193B-18	Sequence 18, Appl
33	37.8	10.6	1582	3 US-08-545-196B-10	Sequence 10, Appl
34	37.8	10.6	1582	3 US-08-545-196B-12	Sequence 12, Appl
35	37.6	10.5	2107	4 US-09-180-852-1	Sequence 1, Appl
36	37.6	10.5	2422	1 US-07-867-106-5	Sequence 5, Appl
37	37.6	10.5	3138	1 US-07-867-106-4	Sequence 4, Appl
38	37.6	10.5	5852	1 US-07-867-106-2	Sequence 2, Appl
39	37.6	10.5	6078	4 US-09-173-914-1	Sequence 1, Appl
40	37.4	10.4	477	4 US-08-887-534A-81	Sequence 81, Appl
41	37.4	10.4	2445	6 5215909-9	Patent No. 5215909
42	37.4	10.4	2674	4 US-09-817-180-1	Sequence 1, Appl
43	37.4	10.4	3761	4 US-08-890-865A-2	Sequence 2, Appl
44	37.4	10.4	10223	4 US-08-961-527-73	Sequence 73, Appl
45	37.2	10.4	144	1 US-08-702-344-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-340-820-24/C
Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
APPLICANT: NARUO, Ken-ichi
APPLICANT: SEKO, Chisako
APPLICANT: KUROKAWA, Tsutomu
APPLICANT: KONDO, Tatsuya
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
NUMBER OF INVENTION: PRODUCTION
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESS: CUSHMAN
STREET: 130 Water Street
City: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/835,713
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, David G.
REGISTRATION NUMBER: 27026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

Db 1116 TTTTACTTACTTACTTACTTCCGAAGAA 1089

RESULT 6

PCT-US92-06300-1/C

Sequence 1, Application PC/TUS9206300
GENERAL INFORMATION:
APPLICANT: Hurwitz, David R
APPLICANT: Nathan, Margret
APPLICANT: Shaul, Moshe
TITLE OF INVENTION: Transgenic Protein Production
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer, Inc.
STREET: 500 Virginia Ave., Bldg. 3A
CITY: Ft. Washington
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19034

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06300
FILING DATE: 19920730
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Goodman, Rosanne
REGISTRATION NUMBER: 52,534
REFERENCE/DOCKET NUMBER: A0856-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 962-4130
TELEFAX: (215) 962-4107

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 19557 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO
PUBLICATION INFORMATION:
AUTHORS: Minghetti, P P
AUTHORS: Ruffner, D E
AUTHORS: Kuang, W.-J.
AUTHORS: Denison, O E
AUTHORS: Hawkins, J W
AUTHORS: Beattie, W G
AUTHORS: Dugalczyk, A

TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE
TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22
TITLE: OF CHROMOSOME 4
JOURNAL: J. Biol. Chem.
VOLUME: 261
PAGES: 6747-6757
DATE: 1986
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002

PCT-US92-06300-1

Query Match 11.4%; Score 40.8; DB 5; Length 19557;
Best Local Similarity 52.3%; Pred. No. 2.3;
Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGATGATTTTATCTCTCTTTTGTGATAATTTTAAATCAT 61
Db 423 ATATCATATATATGATATATATTTCTATTTTCTTTTCTTTTCTTTT 364
QY 62 TTTTGGCCCAATGATATATAAATTTGATATAATTTTGTGATATGCTTTT 121
|||||

Db 363 TTTTATTTTATATCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 304
QY 122 TAGTTCGGGTTTGAGAAAGGCTTTCGACTTTCGAATGACATATAT 173
Db 303 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 252

RESULT 7

US-09-073-569-1/C

Sequence 1, Application US/09073569
Patent No. 6084088

GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Grossmann, Angelika
TITLE OF INVENTION: NOVEL TUMOR ANTIGENS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,569
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 97-14
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1733 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 34...1344
OTHER INFORMATION:
US-09-073-569-1

Query Match 11.3%; Score 40.6; DB 3; Length 1733;
Best Local Similarity 56.3%; Pred. No. 2.5;
Matches 76; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 20 TATTTTATCTCTCTTTTGTGATAATTTTAAATTTTGGCCCAATGATA 79
Db 1715 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1656
QY 80 TATATAATTTTGATATAATATATATGATATCTTTTGTGCGGTTGAGAA 139
Db 1655 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1596
QY 140 AGGCTTTCGACTTC 154
Db 1595 AGGCTCTGATATATC 1581
|||||

RESULT 8

5231168-1
; Patent No. 5231168
; APPLICANT: DEIGIEL, MORTEN, BORRE, MARTIN, JENSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 1:
; LENGTH: 3095
5231168-1

Query Match 11.3%; Score 40.6; DB 6; Length 3095;
Best Local Similarity 62.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAGATATTTTATTCCTCTTTTGTGATAATTTAAATCATTT 63
DB 2724 AAGAGAGAAAAAATTTTATATCATTTTTCATTAATATTAATTA 2783
QY 64 TTTTGGCCCATGATATTAATAATTTGGATAAATAATTAAT 106
DB 2784 TATATATTAATATCTGATAAATATTTATGTAATATTAAT 2826

RESULT 9

US-08-078-090-3
; Sequence 3, Application US/08078090
; Patent No. 5739407
; GENERAL INFORMATION:
; APPLICANT: BERGSTROM, SVEN
; APPLICANT: HERNELL, OLIE
; APPLICANT: LOENNERDAL, BO
; APPLICANT: HALMARSSON, KARIN
; APPLICANT: HANSSON, LENNART
; APPLICANT: TOERNELL, JAN
; APPLICANT: STROMQUIST, MATS
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING
; TITLE OF INVENTION: IT AND USE THEREOF
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy, disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/078,090
; FILING DATE: 19930618
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK92/00236
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK91/00233
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BERGSTROM2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)628-5197
; TELEFAX: (202)737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

LENGTH: 10607 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845
LOCATION: ..6886, 7991..8521, 9440..9445)
US-08-078-090-3

Query Match 11.3%; Score 40.4; DB 1; Length 10607;
Best Local Similarity 56.9%; Pred. No. 2.8;
Matches 74; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 60 ATTTTGTTCGCCATGATATTAATAATTTGGATAAATATATTTGATGATTCGTT 119
DB 923 ATTTATGATATAGAAATGAGTATTAATAACTAGTATATGATATATATGAT 982
QY 120 TTTAGTTCGGCTTGAGAAAAAGGCTTCGACTTCGAAAGCGACATGATATG 179
DB 983 ATTAGTATATTAATATACATTTGATATATATATGATATGATATTAATTA 1042
QY 180 GGACTAGCT 189
DB 1043 GTAGATTAGT 1052

RESULT 10

US-08-749-522-2
; Sequence 2, Application US/08749522
; Patent No. 6096950
; GENERAL INFORMATION:
; APPLICANT: John, Malyakal
; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,522
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 670513.90244
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-2

Query Match 11.2%; Score 40; DB 3; Length 2755;
Best Local Similarity 48.3%; Pred. No. 3.3;
Matches 171; Conservative 0; Mismatches 180; Indels 3; Gaps 2;

QY 6 AAAAAAAAAAGATATTTTATTCCTCTTTTGTGATAAATTAATCATTTT 65

Db 2046 ATATACTAGTAAGTATATATTTGACCTTTTAAACGACCTTATGTTGCTATTTTCG 2105
Oy 66 TTGGCCCAATGATATATAAATTTGATTAATATATATATGATTTTCGTTTTCGT 125
Db 2106 TTAATACCAATCAATATAATTTTATTTATTTATTAATATATGCTACTACAGATGTGG 2165
Oy 126 TCGGCTTGAGAAAAGGCTTCGA-CTTTCGAAAGTCGACGATATATAGATTGGAGC 184
Db 2166 TGAAGTAAATATTTTAAACAATATATTTTGAATAATGATTAATATGAGGTT 2225
Oy 185 TAGGTGAGCTTTGACATTTGTATGATGTTGTTTATTTATTTGTCGACACTATA 244
Db 2226 TTGGTGAATAGTATATATATTTTACAAATTTATTTATGATGTTCAAAATCTATC 2285
Oy 245 AACCTTAATAGGCTTTCTATAGGCCCAATTAATATGATTTATACAAAGTCACACT 304
Db 2286 ATGTGATATTTG--TACTATATTTCTATATTAATGATTAACCTTATATAAGTATCTANT 2343
Oy 305 TTATCTGCTTTTGTGATCCGAAAGCAATATACAAATTTGCAATACCAACACAG 358
Db 2344 TTAGTTATGTTGTTGATGATGATGATACCAAAATTTATTAATATATATAG 2397

RESULT 11

US-08-341-568-3/C
Sequence 3, Application US/08341568
Patent No. 5661021

GENERAL INFORMATION:

APPLICANT: Buchert, Johanna
APPLICANT: Silka-aho, Matti
APPLICANT: Vilkarl, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
NUMBER OF INVENTION: 7
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA
COUNTRY: US
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/341,568
FILING DATE: 22-NOV-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: Trichoderma reesei
STRAIN: QM9414

US-08-341-568-3

Query Match

Best Local Similarity 11.0%; Score 39.4; DB 1; Length 289;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Oy 20 TATTTTATTCCTTTTGTGATTAATTTAAATCAATTTTGGCCCAATGATA 79
Db 277 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTGATA 218
Oy 80 TATTAATTTGATTAATA 100
Db 217 TATATATATATTCATCAAAA 197

RESULT 12

US-08-911-020-3/C
Sequence 3, Application US/08911020
Patent No. 5854047

GENERAL INFORMATION:

APPLICANT: Buchert, Johanna
APPLICANT: Silka-aho, Matti
APPLICANT: Vilkarl, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA
COUNTRY: US
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,020
FILING DATE: 13-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/341,568
FILING DATE: 22-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: Trichoderma reesei
STRAIN: QM9414
US-08-911-020-3

Query Match

Best Local Similarity 11.0%; Score 39.4; DB 2; Length 289;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;


```

OY      6  AAAAAAGATGAAAGTATTTTATTCCTCTCTTTTGTGATATTTTAAATCATTTT 65
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1513  AAAAAATATTTTGAATTTTATTTTATTTTATTTTAAATTAATTTTACATCTTTTA 1572
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      66  TTGCCCAATGATATATTTTAAATTTTGATTAATATATTTGATATTCGTTTGTAGT 125
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1573  ATACGTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1632
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      126  TCGGGTTTGAGAAAA 140
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1633  TAATATTTTAAAAAAA 1647
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: June 11, 2003, 01:53:51
 Job time : 18.0513 secs

Run on: June 11, 2003, 01:50:19 ; Search time 51 3843 seconds

9750.738 Million cell updates/sec

```

Title: US-09-502-426B-1_COPY_6111_6468
Perfect score: 358
Sequence: 1 aaaaaaaaaaaaaaaatctttttttt

```

Scoring table: IDENTITY_NUC

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Maximum DB seq length: 20000000000

Processing: Minimum Match 08
Maximum Match 100%

Database : Published Applications wa.

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

pred. no. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50.8	14.2	13606	9	US-10-239-676-165	Sequence 165, App
2	49.8	13.9	8866	9	US-10-239-676-140	Sequence 140, App
3	49.2	13.7	11036	9	US-10-239-676-117	Sequence 117, App
4	49	13.7	12405	9	US-10-239-676-36	Sequence 36, App
5	48.4	13.5	5951	9	US-10-239-676-115	Sequence 115, App
6	48.4	13.5	7823	9	US-10-239-676-198	Sequence 198, App
7	48	13.4	6327	9	US-10-239-676-151	Sequence 151, App
8	47.8	13.4	516	10	US-09-960-352-5785	Sequence 5785, App
9	46.8	13.1	5510	10	US-10-239-676-170	Sequence 170, App
10	46.8	13.1	9515	9	US-10-239-676-160	Sequence 160, App
11	46.6	13.0	11735	9	US-10-239-676-33	Sequence 33, App
12	46.4	13.0	6306	9	US-10-239-676-223	Sequence 223, App
13	46.2	12.9	6397	9	US-10-239-676-107	Sequence 107, App
14	45.4	12.7	12465	9	US-10-239-676-31	Sequence 31, App
15	45.4	12.7	15732	9	US-10-239-676-95	Sequence 95, App
16	45	12.6	189	9	US-10-102-806-401	Sequence 401, App
17	45	12.6	11836	9	US-10-239-676-102	Sequence 102, App
18	44.6	12.5	277	10	US-09-960-352-12673	Sequence 12673, App
19	44.6	12.5	681	10	US-09-960-352-8304-26	Sequence 26, App

20	44.6	12.5	7588	9	US-10-239-676-191	Sequence 191, App
21	44.6	12.5	11735	9	US-10-239-676-34	Sequence 34, App
22	44.6	12.5	12405	9	US-10-239-676-35	Sequence 35, App
23	44.4	12.4	5327	9	US-10-239-676-61	Sequence 61, App
24	44.4	12.4	6620	9	US-10-239-676-161	Sequence 161, App
25	44.4	12.4	7544	9	US-10-239-676-6	Sequence 6, App
26	44.4	12.4	9254	9	US-10-239-676-86	Sequence 86, App
27	44.2	12.3	6306	9	US-10-239-676-130	Sequence 130, App
28	43.8	12.3	6397	9	US-10-239-676-108	Sequence 108, App
29	43.8	12.3	277	10	US-09-960-352-5094	Sequence 5094, A
30	43.8	12.2	446	10	US-09-960-352-3400	Sequence 3400, A
31	43.8	12.2	447	10	US-09-960-352-759	Sequence 759, A
32	43.6	12.2	413	10	US-09-960-352-2919	Sequence 2919, A
33	43.6	12.2	6030	9	US-10-239-676-164	Sequence 164, App
34	43.6	12.2	7089	9	US-10-239-676-67	Sequence 67, App
35	43.6	12.2	7657	9	US-10-239-676-185	Sequence 185, App
36	43.6	12.2	10619	9	US-10-239-676-2	Sequence 2, App
37	43.4	12.1	7147	9	US-09-918-995-7147	Sequence 7147, App
38	43.4	12.1	5610	9	US-10-239-676-169	Sequence 169, App
39	43.4	12.1	5689	9	US-10-239-676-90	Sequence 90, App
40	43.4	12.1	6534	9	US-10-239-676-111	Sequence 111, App
41	43.4	12.1	11836	9	US-10-239-676-103	Sequence 103, App
42	43.4	12.1	15649	9	US-10-239-676-103	Sequence 103, App
43	43.2	12.1	414	10	US-09-960-352-6528	Sequence 6528, A
44	43.2	12.1	570	9	US-10-198-846-2709	Sequence 2709, App
45	43.2	12.1	2933	9	US-10-028-072-345	Sequence 345, App

ALIGNMENT

RESULT 1
US-10-239-676-165

Publication No. US20030082609A1

APPLICANT: PIEPENBROCK, christia

FILE REFERENCE: 5013.1003

CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03000

DE 10019173.8
DE 10032530.7

DE 10043826.1
PRIOR FILING DATE: 2001-04-03

2000-04-07
2000-06-30

NUMBER OF SEQ ID NOS: 22
SEQ ID NO 165

TYPE: DNA

OTHER INFORMATION: chemically

NAME/KEY: unsure
LOCATION: /A649 763A 76E7)

Quorum Watch

Best Local Similarity 52.3%; Pred. No. 0.83;
Matches 112; Conservative 0; Mismatches 103.

Y
20 TATTTCCTCCTCTTTTTCGTAATTTAAATCATTTTTTGGCCCAATGATA 7

11066


```

RESULT 5
US-10-239-676-115
: Sequence 115, Application US/10239676
: Publication No. US20030082609A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
: FILE REFERENCE: 5013.1003
: CURRENT APPLICATION NUMBER: US/10/239.676
: CURRENT FILING DATE: 2002-05-24
: PRIOR APPLICATION NUMBER: PCT/EP01/03968
: DE 10019058.8
: DE 10019173.8
: DE 10032529.7
: DE 10043826.1
: PRIOR FILING DATE: 2001-04-06
: 2000-04-06
: 2000-04-07
: 2000-06-30
: 2000-09-01
: NUMBER OF SEQ ID NOS: 228
: SEQ ID NO 115
: LENGTH: 5951
: TYPE: DNA
:

```

Query Match	13.5%	Score	48.4	DB	9	length	7823
Best Local Similarity	46.9%	Pred. No.	2.1				
Matches	151	Conservative	0	Mismatches	171	Indels	0
				Gaps	0		
QY	4	AAAAAAAAAGAGCAAGTATTTTATCTCTCTCTTTTGTGATAAATTTAAACATT	63				
Db	32	AAATATGTTATGTTTATGTAATAATTTTTTTTTTTTAAATAGTATATGTA	91				

[illegible]

DE 10019058.8

DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 95
LENGTH: 15732
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-95

Query Match 12.7%; Score 45.4; DB 9; Length 15732;
Best Local Similarity 49.8%; Pred. No. 10;
Matches 143; Conservative 0; Mismatches 141; Indels 3; Gaps 1;
QY 14 TGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCATTTTGGCCA 73
DB 10254 TTAAGTTTGTAGTATTAATTTTAAAGAAATTAATTTTATTTGATTT 10313
QY 74 ATGATATATAAAATTTGATTAATTAATTAATTCGTTTATTTAGTGGGTTT 133
DB 10314 TTTTATTAATTTTAAATTTTAAATTTTAAATTTTATTTTATTTTGA 10373
QY 134 GAGAAAGGTTTCGACTTCGAAAGTGACGATGATATAGATTGGAGCTAGGTTGAG 193
DB 10374 TGGTTTGTGTTTGTGTTTATTTTAAAGTTTATGTTAAAGTTAGATTAT 10433
QY 194 TCTTGGACATTTGATG--ATGTTGTGATTATTAAGTTCGACACTATAACCTT 250
DB 10434 TGATTTGAGAGTTTGTGTAAGTGAATGTTTAAATTTTAAATTTTATTTATT 10493
QY 251 AATGGGCTTCTATAGGCCCAATTATATAGATTATTAACAAGT 297
DB 10494 ATATTTTGTATTAATGAATGATATGTTGTGTTTATTAATTAAT 10540

Search completed: June 11, 2003, 08:07:42
Job time : 52.3843 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 860.091 Seconds

(without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcataataccaacaagaag 358

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49587208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

```

1: Pending_Patents_NA_Main:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US099E_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*

```

No.	Score	Query	%	Match	Length	DB	ID	Description
1	358	100.0	6888	1	PCT-US00-03820-1			Sequence 1, Appl
2	358	100.0	6888	19	US-09-502-426-1			Sequence 1, Appl
3	356.4	99.6	745	19	US-09-505-532-9934			Sequence 9934, Ap
4	356.4	99.6	745	31	US-09-819-091A-9934			Sequence 9934, Ap
5	356.4	99.6	84196	27	US-09-692-412-94			Sequence 94, Appl
6	356.4	99.6	84196	31	US-09-803-736-1074			Sequence 1074, Ap
7	110.4	30.8	566	25	US-09-654-617-133308			Sequence 133308, Ap
8	110.4	30.8	566	27	US-09-684-016-133308			Sequence 133308, Ap
9	55.8	15.6	495	38	US-10-021-323-17366			Sequence 17366, A
10	55.8	15.6	495	69	US-09-654-617-133963			Sequence 133963, Ap
11	54.6	15.3	284	25	US-09-654-617-421464			Sequence 421464, A
12	54.6	15.3	284	27	US-09-684-016-339963			Sequence 339963, Ap
13	54.6	15.3	284	58	US-09-619-643-14496			Sequence 14496, A
14	53.8	15.0	286	33	US-09-865-419A-12520			Sequence 12520, A
15	53.2	14.9	572	33	US-10-172-086-51			Sequence 51, Appl
16	53.2	14.9	14147	41	US-09-654-617-421464			Sequence 421464, A
17	53	14.8	384	27	US-09-684-016-421464			Sequence 421464, A
18	53	14.8	384	25	US-09-614-150-25744			Sequence 25744, A
19	52.6	14.7	21231	63	US-09-614-150-25744			Sequence 25744, A
20	52.6	14.7	21231	63	US-09-614-150-25744			Sequence 25744, A
21	52.6	14.7	21241	63	US-60-191-637-25874			Sequence 25874, A

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Sequence	20483, A
Sequence	10689, A
Sequence	29, APL
Sequence	12650, A
Sequence	18907, A
Sequence	556, APP
Sequence	180781, A
Sequence	16073, A
Sequence	180181, A
Sequence	339755, A
Sequence	339755, A
Sequence	5695, APP
Sequence	367844, A
Sequence	367844, A
Sequence	35283, A
Sequence	165, APP
Sequence	61929, A
Sequence	79531, A
Sequence	51282, A
Sequence	31950, A
Sequence	319872, A
Sequence	339872, A
Sequence	6045, APP
Sequence	28359, A

ORGANISM: Arabidopsis sp.
ITS-09-502-426-1

```

OY      301  AACTTTTACTCGTTTGTGATCCGANGCATATACAAATGTCTAAATACCAACACAG 358
        |||||||
Db      6411  AACTTTTACTCGTTTGTGATCCGANGCATATACAAATGTCTAAATACCAACACAG 6468

        RESULT 2
        ; Sequence 1, Application US/09502426
        ;
        ; GENERAL INFORMATION:
        ; APPLICANT: AZIPROZ, Ricardo
        ; APPLICANT: CHOE, Sunghwa
        ; APPLICANT: FELDMANN, Kenneth
        ; TITLE OF INVENTION: DMPA POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
        ; FILE REFERENCE: 2225-0001
        ; CURRENT APPLICATION NUMBER: US/09/502,426
        ; CURRENT FILING DATE: 2000-02-11
        ; EARLIER APPLICATION NUMBER: 60/119,657
        ; EARLIER FILING DATE: 1999-02-11
        ; EARLIER APPLICATION NUMBER: 60/119,658
        ; EARLIER FILING DATE: 1999-02-11
        ; NUMBER OF SEQ ID NOS: 18
        ; SOFTWARE: PatentIn Ver. 2.0
        ; SEQ ID NO 1
        ; LENGTH: 6888
        ; TYPE: DNA
        ; ORGANISM: Arabidopsis sp.
        ; US-09-502-426-1

```

```

Query Match      100.0%; Score 358; DB 19; Length 6888;
Best Local Similarity 100.0%; Pred. No. 9 2e-44;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      AAAAAAAAAAAGATGAAAGATATTTTATATCTCTCTCTTTTATTTTGTATATTTAAATCA 60
Db      6111     AAAAAAAAAAAGATGAAGATATTTTATATCTCTCTTTTATTTTGTATATTTAAATCA 6170

QY      61      TTTTGTGGCCATGATATATAAAATTTGATAAATAATATATATGATATTCGTTT 120
Db      6171     TTTTGTGGCCATGATATATAAAATTTGATAAATAATATATATGATATTCGTTT 6230

QY      121     TTACTTCGGGTTTGAGAAAAGGGTTTGACCTTTGGAAAGTCGACAGCTATATAGATTGG 180
Db      6231     TTACTTCGGGTTTGAGAAAAGGGTTTGACCTTTGGAAAGTCGACAGCTATATAGATTGG 6290

QY      181     GAGCTAGGTGAGTCTTTGACATTTGTATGAGATGTTGATTATTTAGTGGACACT 240
Db      6291     GAGCTAGGTGAGTCTTTGACATTTGTATGAGATGTTGATTATTTAGTGGACACT 6350

QY      241     ATTAACCTTAATAGGCGTTCTCTATAAGGCCCAATATATACGTTTAAACAAAGTAC 300
Db      6351     ATTAACCTTAATAGGCGTTCTCTATAAGGCCCAATATATATTCGATTTAAACAAAGTAC 6410

QY      301     AACCTTACTTCGTTTGTATCCGAAAGCAATTAACAAATTTGTCAATACCAACACAG 358
Db      6411     AACCTTACTTCGTTTGTATCCGAAAGCAATTAACAAATTTGTCAATACCAACACAG 6468

RESULT 3
US-09-505-532-9934/C
/ Sequence 9934, Application US/0950532
/ GENERAL INFORMATION:
/ APPLICANT:      Cao, Yongwei
/ APPLICANT:      Timberlake, William E.
/ TITLE OF INVENTION:  Plant Genome Sequences and Uses Thereof
/ FILE REFERENCE:    38-10(15478)B
/ CURRENT FILING DATE: 2000-02-16
/ PRIOR APPLICATION NUMBER:  USN 60/120,645 1999-02-18;  USN 09/443,025
/ PRIOR FILING DATE:   USN 60/120,645 1999-02-18;  USN 09/443,025 1999-11-1
/ NUMBER OF SEQ ID NOS: 51470
/ SEQ ID NO 9934
/ LENGTH: 745

```


PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ ID NOS: 1582
SEQ ID NO 1074
LENGTH: 84196
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-803-736-1074

Query Match
Best Local Similarity 99.6%; Score 356.4; DB 31; Length 84196;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 26092 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 26033
QY 61 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 120
DB 26032 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 25973
QY 121 TTAGTTCGGGTTGAGAAAGGTTTGACCTTCGAAAGTGACGATGATTAATGATGG 180
DB 25972 TTAGTTCGGGTTGAGAAAGGTTTGACCTTCGAAAGTGACGATGATTAATGATGG 25913
QY 181 GACCTAGGTTGACCTTTGACATTTGATGATGTTGATTAATTAATGATGACACT 240
DB 25912 GACCTAGGTTGACCTTTGACATTTGATGATGTTGATTAATTAATGATGACACT 25853
QY 241 ATTAACCTTAATAGGCTTCTATAGGCCCAATTAATTAATTAATTAATTAATGAC 300
DB 25852 ATTAACCTTAATAGGCTTCTATAGGCCCAATTAATTAATTAATTAATTAATGAC 25793
QY 301 AACTTCTACTCGTTTGTGATCCGACCAATTAACAAATTCGCAATCAACCAAG 358
DB 25792 AACTTCTACTCGTTTGTGATCCGACCAATTAACAAATTCGCAATCAACCAAG 25735

RESULT 7
US-09-654-617-133308/C
Sequence 133308, Application US/09654617
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
TITLE OF INVENTION: Annotated Plant Genes
FILE REFERENCE: 38-21(15097)D
CURRENT APPLICATION NUMBER: US/09/654,617
CURRENT FILING DATE: 2000-09-05
NUMBER OF SEQ ID NOS: 463173
SEQ ID NO 133308
LENGTH: 566
TYPE: DNA
ORGANISM: Arabidopsis thaliana
OTHER INFORMATION: unsure at all n locations
US-09-654-617-133308

Query Match
Best Local Similarity 90.8%; Score 110.4; DB 25; Length 566;
Matches 128; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 232 AAAAAAAAAAGATGAAGATATTTATCTCTCTCTTTTGTGATTAATTAATCA 174
QY 61 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 120
DB 173 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 114
QY 121 TTAGTTCGGGTTGAGAAAG 141
DB 113 TTTNCCCAAAAAAAAAAAG 93

RESULT 8
US-09-684-016-133308/C
Sequence 133308, Application US/09684016
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: Annotated Plant Genes
FILE REFERENCE: 38-21(15097)D
CURRENT APPLICATION NUMBER: US/09/684,016
CURRENT FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/654,617
PRIOR FILING DATE: 2000-09-05
NUMBER OF SEQ ID NOS: 463173
SEQ ID NO 133308
LENGTH: 566
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(566)
OTHER INFORMATION: unsure at all n locations
US-09-684-016-133308

Query Match
Best Local Similarity 90.8%; Score 110.4; DB 27; Length 566;
Matches 128; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 232 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 174
QY 61 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 120
DB 173 TTTTGTGCCCCATGATATATAAATTTGATTAATTAATTAATTCGTTT 114
QY 121 TTAGTTCGGGTTGAGAAAG 141
DB 113 TTTNCCCAAAAAAAAAAAG 93

RESULT 9
US-10-021-323-17366/C
Sequence 17366, Application US/10021323
GENERAL INFORMATION:
APPLICANT: Deikman, Jill
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(52274)B
CURRENT APPLICATION NUMBER: US/10/021,323
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: US 60/255, 619
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 17880
SEQ ID NO 17366
LENGTH: 495
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: LIB3830-001-Q1-N6-B2
US-10-021-323-17366

Query Match
Best Local Similarity 51.4%; Score 55.8; DB 38; Length 495;
Matches 129; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGATGAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 61
DB 257 ATAAAGTTAAAGAGGTGACGACATTTAACTTTTTCATGAAAGAAAAAAT 198


```

; RESULT 11
; US-09-654-617-339963/C
; Sequence 339963, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15087)D
; CURRENT APPLICATION NUMBER: US/09/654, 617
; CURRENT FILING DATE: 2000-09-05

```

```

RESULT 12
US-09-684-016-339963/C
; Sequence 339963, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684, 016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 339963
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Zea mays
US-09-684-016-339963

Query Match          15.3%; Score 54.6; DB 27; Length 284;
Best Local Similarity 56.4%; Pred. No. 54;
Matches 102; Conservative 0; Mismatches 79; Indels 0; Gaps 0

QY      6 AAAAAAGATGAAGAATTTTATTCCTCTCTTTTATTTTTCGTAATTTTAATCAATTTT 65
Db      276 AATAAAAAATTAATTTTATTTTATTTTATTTTATTTTATTAATAATTAATTA 217

QY      66 TTGGCCATGATATATATAAAAAATTTGGATAAATATATATGATATTCGTTTTTAC 125
Db      216 TTATTTATTTATTTATTTATTTAATTTTATTTTATTTATTTATTTATTTATTT 157

QY      126 TCGGGTTTGAGAAAAAGGTTTCGACTTTCGAAAGTCAGCAGATATATAGATTGGAGCT 185
Db      156 TTTTATTTTAAATTAATTAATAAATATATAATATTAATAAATATAAATATAAAT 97

QY      186 A 186
Db      96 A 96

RESULT 13
US-60-145-485-6295/C
; Sequence 6295, Application US/60145485
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine

```

```

: APPLICANT: Deng, Molian
: APPLICANT: Fisher, Dane K.
: APPLICANT: Miller, Philip W.
: TITLE OF INVENTION: NOCLETIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: PLANTS
: FILE REFERENCE: 38-21(51229)A
: CURRENT APPLICATION NUMBER: US/60/145,485
: CURRENT FILING DATE: 1999-07-23
: NUMBER OF SEQ ID NOS: 9366
: SEQ ID NO 6295
: LENGTH: 284
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LJB3150-079-P2-N2-B2
US-60-145-485-6295

```

Query Match	15.38;	Score 54, 6;	DB 58;	Length 284;
Best Local Similarity	56.48;	Pred. NO. 54;		
Matches 102; Conservative	0;	Mismatches 79;	Indels 0;	Gaps 0

QY 6 AAAAAAGATGAAATTTTATTTATTCCTCTTTTTTTTTTGGATTAATTTAAATCATTTT 65
 |||||
 DB 276 AATAAAATTAATTTTATTTTATTTTATTAATTTTAAATTTTAAATTTTAAATTTA 217

0y 66 TTTCGCCATGATAATAAAAATTGGATAAATAATTTATGCGATATCGTTTTTNGT 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 TTTTATTATATATTTATTAAATTTTTTTTATTAAATTAATATATATATATATATATATAT 157

Oy 126 TCGGGTTTGAAGAAAGGCTTCGACTTTCGAAGGTGCACGATATATAGATTGGGACT 185
 |||||
Db 156 TTTTTTTTAAAAATTATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAAT 97

Oy	186 A 186
	I
Db	96 A 96

```

: RESULT 14
: US-09-619-643-14496/c
: Sequence 14496, Application US/09619643
: GENERAL INFORMATION:
: APPLICANT: Fisher, Dane K.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: FILE OF INVENTION: Plants
: FILE REFERENCE: 38-21(51230)B
: CURRENT APPLICATION NUMBER: US/09/619,643
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 32236
: SEQ ID NO 14496
: LENGTH: 286
: TYPE: DNA
: ORGANISM: Zea mays
: OTHER INFORMATION: Clone ID: LTB3150-079-P2-N2-B2
: US-09-619-643-14496

```

Query Match	15.3%	Score 54.6	DB 23	Length 286
Best local similarity	56.4%	Pred. No. 54		
Matches 102; Conservative	0	Mismatches 79	Indels 0	Gaps 0

6 AAAAAGATGAAGTATTTTATTCTCTCTTTTTCATTAATTTAAATCATTTT 65

Db 278 ATAAAAAATTAAATTTTATTTTATTTTATTTTAAATTTTAAATTTA 219

66 TTTGCCCAATGATATATAAAAAATTTGGATAAATAATATATATGGATATTCGTTTTTACT 125

[illegible][illegible]

QY	186 A 186
	-
Db	98 A 98

```

RESULT 15
US-09-865-419A-12520/c
: Sequence 12520, Application US/09865419A
: GENERAL INFORMATION:
: APPLICANT: Conner, Timothy W.
: APPLICANT: Wu, Kunsheng
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plasts
: FILE REFERENCE: 38-21(51935)B
: CURRENT APPLICATION NUMBER: US/09/865,419A
: CURRENT FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/208,063
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ. ID NOS: 54020
: SEQ ID NO 12520

```

```

? TYPE: DNA
? ORGANISM: Zea mays
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (1)..(572)
? OTHER INFORMATION: unsure at all n locations
? OTHER INFORMATION: Clone ID: LIB3236-144-PI-N1-H-
? OS-09-865-419A-12550

```

Query Match	15.0%	Score 53.8;	DB 33;	Length 572;
Best Local Similarity	64.8%	Pred. No. 71;		
Matches 79; Conservative	0;	Mismatches 43;	Indels 0;	Gaps 0

Qy 2 AAAAAAAAAAGATGAAAGTATTTTAATCTCTCTTTTTTTTTTGATAATTATAATCAT 61
| | | | |
Db 392 AAAAAAAAAATTTTAAATTTTTTTTTTTTTANTTTTTTTTTTTTAAAAATTTAAAATTT 33

QY	62 TTTTTCCTCCCAATGATATAAAATTGGATAAATAATPTATTGGATATTCGTATT	121
	- - - -	
Db	332 TTTTTTTTTTTTTTAATAAAATTTTTTTTTTAAAAAAAATTTTTTTTTTTTTTTTTT	273

QY	122	TA	123
Db	272	TA	271

Search completed: June 11, 2003, 05:00:52
Job time : 871.091 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:22:54 ; Search time 265.216 Seconds
(without alignments)
8960.952 Million cell updates/sec

Title: US-09-502-426B-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaaagatgaagf.....tgcacatcacaacacaaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 7602234 seqs, 3319262570 residues

Total number of hits satisfying chosen parameters: 15204468

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	358	100.0	6888	US-09-502-426B-1	Sequence 1, Appl
2	358	100.0	6888	US-09-502-426B-1	Sequence 1, Appl
3	58	16.2	5449	US-10-311-455-1614	Sequence 1614, Ap
4	58	16.2	5449	US-10-311-455-1614	Sequence 1614, Ap
5	56.4	15.8	5926	US-10-311-455-1625	Sequence 1196, Ap
6	55	15.4	7351	US-10-311-455-1625	Sequence 1625, Ap
7	55	15.4	7351	US-10-311-455-1625	Sequence 1625, Ap
8	54.8	15.3	20486	US-10-240-485-164	Sequence 164, Appl
9	54.2	15.1	3673778	US-10-312-841-2	Sequence 2, Appl
10	54	15.1	2814	US-10-311-455-2100	Sequence 2100, Ap
11	53.2	14.9	14147	US-10-240-453-39	Sequence 329, App
12	53.2	14.9	14147	US-10-311-506-39	Sequence 39, Appl
13	53.2	14.9	14147	US-10-311-507-39	Sequence 39, Appl
14	53.2	14.9	14147	US-10-311-507-39	Sequence 39, Appl
15	52.8	14.8	384	US-09-837-606A-39	Sequence 2063, Ap
16	52.8	14.7	5678	US-10-311-455-2063	Sequence 1111, Ap
17	52.2	14.6	6418	US-10-311-455-295	Sequence 295, App
18	51.8	14.5	16750	US-10-311-455-494	Sequence 494, App
19	51.6	14.4	17131	US-10-311-455-945	Sequence 945, App
20	51.2	14.3	3673778	US-10-311-455-1026	Sequence 1026, App
21	51	14.2	13584	US-10-311-455-588	Sequence 588, Appl
22	50.8	14.2	5311	US-10-311-455-992	Sequence 992, App

23	50.8	14.2	8087	US-10-311-455-715	Sequence 715, App
24	50.8	14.2	13606	US-10-240-453-187	Sequence 187, App
25	50.8	14.2	13606	US-10-311-455-1783	Sequence 1783, App
26	50.6	14.1	5649	US-10-257-166-90	Sequence 90, Appl
27	50.6	14.1	5649	US-10-311-455-822	Sequence 822, App
28	50.4	13.9	5163	US-10-311-455-1221	Sequence 1221, App
29	49.8	13.9	6759	US-10-311-455-711	Sequence 711, App
30	49.8	13.9	8866	US-10-240-453-154	Sequence 154, App
31	49.8	13.9	113515	US-10-311-455-2148	Sequence 2148, App
32	49.6	13.9	9347	US-10-204-708-36	Sequence 36, Appl
33	49.6	13.9	11422	US-10-257-166-17	Sequence 17, Appl
34	49.6	13.9	11422	US-10-311-455-191	Sequence 191, App
35	49.6	13.9	13427	US-10-311-455-1899	Sequence 1899, App
36	49.4	13.8	7238	US-10-311-455-421	Sequence 421, App
37	49.4	13.8	8201	US-10-240-452-27	Sequence 27, Appl
38	49.4	13.8	8201	US-10-311-455-279	Sequence 279, App
39	49.4	13.8	9964	US-10-311-455-71	Sequence 71, Appl
40	49.2	13.7	432	US-09-837-606A-67427	Sequence 67427, A
41	49.2	13.7	6464	US-10-311-455-487	Sequence 487, App
42	49.2	13.7	11036	US-10-240-453-137	Sequence 137, App
43	49	13.7	6032	US-10-240-454-6	Sequence 6, Appl
44	49	13.7	6494	US-10-311-455-1366	Sequence 1366, App
45	49	13.7	12405	US-10-240-453-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1	US-09-502-426B-1	Sequence 1, Application US/09502426B
GENERAL INFORMATION:		
APPLICANT:	Asplioz, Ricardo	
APPLICANT:	Choe, Sungjwa	
APPLICANT:	Feldmann, Kenneth A.	
TITLE OF INVENTION:	DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF	
FILE REFERENCE:	11696-070001	
CURRENT APPLICATION NUMBER:	US/09/502,426B	
CURRENT FILING DATE:	2000-02-11	
PRIOR APPLICATION NUMBER:	US 60/119,657	
PRIOR FILING DATE:	1999-02-11	
PRIOR APPLICATION NUMBER:	US 60/119,658	
PRIOR FILING DATE:	1999-02-11	
NUMBER OF SEQ ID NOS:	30	
SOFTWARE:	FastSeq For Windows Version 4.0	
SEQ ID NO 1		
LENGTH:	6888	
TYPE:	DNA	
ORGANISM:	Arabidopsis thaliana	
US-09-502-426B-1		
Query Match	100.0%; Score 358; DB 5; Length 6888;	
Best Local Similarity	100.0%; Pred. No. 7e-53;	
Matches 358; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
1	AAAAAAAAAAGAGTAAGTATTTTATCTCTCTTTTGTGATTAATCA 60	
6111	AAAAAAAAAAGAGTAAGTATTTTATCTCTCTTTTGTGATTAATCA 6170	
61	TTTTTTTGGCCATGATATATAAATTTGATATAATATTGATTCGTTT 120	
6171	TTTTTTTGGCCATGATATATAAATTTGATATAATATTGATTCGTTT 6230	
121	TTAGTGGGTTTGAAGAGGTTTGACATTTGCAAGTGCAGATGATATGATTGG 180	
TTAGTGGGTTTGAAGAGGTTTGACATTTGCAAGTGCAGATGATATGATTGG 6290		
181	GAGCTAGGTTGAGCTTTGACATTTGATGATGTTGATTTAGTGGACACT 240	
6291	GAGCTAGGTTGAGCTTTGACATTTGATGATGTTGATTTAGTGGACACT 6350	
241	ATTAACTTAATAGGCTTTCTTAAGCCCAATTAATTAGCATTAACAAGTAC 300	

Db 6351 ATTAACCTTAATGGGCTTCTATTAAGCCCAATTATATACATTATAACAAGTGAC 6410
QY 301 AACTTTACTTCGTTTTTGATCCGACCAATTAACAATTTGCAATACCAACAG 358
Db 6411 AACTTTACTTCGTTTTTGATCCGACCAATTAACAATTTGCAATACCAACAG 6468

RESULT 2

US-09-502-426a-1
; Sequence 1, Application US/09502426a
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426a
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
US-09-502-426a-1

Query Match 100.0%; Score 358; DB 7; Length 6888;
Best Local Similarity 100.0%; Pred. No. 7e-53;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 60
Db 6111 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 6170
QY 61 TTTTTCGCCCAATGATATATAATAAATTTGGATTAATATATTTGATTTTCGTTT 120
Db 6171 TTTTTCGCCCAATGATATATAATAAATTTGGATTAATATATTTGATTTTCGTTT 6230
QY 121 TTGTCGGGTTTGAGAAAGGTTTGCACCTTGGAAAGTGACGATATATGATTGG 180
Db 6231 TTGTCGGGTTTGAGAAAGGTTTGCACCTTGGAAAGTGACGATATATGATTGG 6290
QY 181 GAGCTAGGTGAGCTTTTGACATTTGATTTGATTTGATTTAGTGTGACACT 240
Db 6291 GAGCTAGGTGAGCTTTTGACATTTGATTTGATTTGATTTAGTGTGACACT 6350
QY 241 ATTAACCTTAATGGGCTTCTATAAGCCCAATTATATACGATTATAACAAGTGAC 300
Db 6351 ATTAACCTTAATGGGCTTCTATAAGCCCAATTATATACGATTATAACAAGTGAC 6410
QY 301 AACTTTACTTCGTTTTTGATCCGACCAATTAACAATTTGCAATACCAACAG 358
Db 6411 AACTTTACTTCGTTTTTGATCCGACCAATTAACAATTTGCAATACCAACAG 6468

RESULT 3

US-10-311-455-1614
; Sequence 1614, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1614
; LENGTH: 5449
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1614

Query Match 16.2%; Score 58; DB 9; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.2;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 63
Db 5256 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 5315
QY 64 TTTTTCGCCCAATGATATATAATAAATTTGGATTAATATATTTGATTTTCGTTT 123
Db 5316 TGTTTAAGCTTTTATATATATTTTATATATTTTATATATTTGATTTTATTT 5375
QY 124 GTTCGGGTTT 133
Db 5376 TTTTGGGTTT 5385

RESULT 4

US-10-311-455-1196
; Sequence 1196, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1196
; LENGTH: 7560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1196

Query Match 16.2%; Score 58; DB 9; Length 7560;
Best Local Similarity 51.6%; Pred. No. 0.19;
Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

QY 2 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 61
Db 2728 ATTAATTAAGCAATTAAGTATTTTATTTTGTGATTAATTTAAATCA 2787
QY 62 TTTTTCGCCCAATGATATATAATAAATTTGGATTAATATATTTGATTTTCGTTT 121
Db 2788 TTTTATGTCGAGCTTATTTTAAATTTTAAAGATATTTTATTTAGTGTGATTT 2847
QY 122 TAGTTCGGGTTTGAGAAAGGTTTGCACCTTGCAGAAAGTGACGATATATATGATTGG 181
Db 2848 TGTTCGATTTATATAAATTTATTTAGCATTTTATTTATTTTAAAAATTAATAT 2907

QY	182	AGCTAGCTGAGTCTTTGGACATTTGATTGGATG----	TGTTGATTATTAGTGCAC	237
Db	2908	TATTATTAATTAATTAATTTGATTAATAGTAATTAAGTGTGATGATGTATTATTTTAATTT	2966	
QY	238	ACATTTAAACCTTAATGGCTTTCTATAGGCCCATTTATTAAGATTATTAACAAGT	297	
Db	2968	ATATTTTGAATTTAATTAATTTTATTTATTAACAATAATTATTAGAAAAATTAATTAATTAAGA	3027	
QY	298	GACAACTT	305	
Db	3028	GATTAATT	3035	

RESULT 5
US-10-31

```
US-10-311-455-1625
; Sequence 1625, Application US/10311455
; GENERAL INFORMATION:
```

```

1  APPLICANT: OLEK, Alexander
2  APPLICANT: PIEPENBROCK, Christian
3  APPLICANT: BERLIN, Kurt
4  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
5  TITLE OF INVENTION: cytosine methylation
6  FILE REFERENCE: 5013.1014
7  CURRENT APPLICATION NUMBER: US/10/311.455
8  CURRENT FILING DATE: 2002-12-16
9  PRIOR APPLICATION NUMBER: PCT/EP01/07537
10 PRIOR FILING DATE: 2001-07-02
11 PRIOR APPLICATION NUMBER: DE 10032529.7
12 PRIOR FILING DATE: 2000-06-30
13 PRIOR APPLICATION NUMBER: DE 10043826.1
14 PRIOR FILING DATE: 2000-09-01
15 NUMBER OF SEQ ID NOS: 2424
16 SEQ ID NO 1625
17 LENGTH: 5926
18 TYPE: DNA
19 ORGANISM: Artificial Sequence
20 FEATURE:
21 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
22 US-10-311-455-1625

```

Query Match	15.88;	Score 56.4;	DB 9;	Length 5926;
Best Local Similarity	48.28;	Pred. No. 0.36;		
Matches 159;	Conservative	0;	Mismatches 171;	Indels 0;
			Gaps	0;

OY	4	AAAAAAAAAGATGAAAGTATTTTATTCCTCCTTTTTTTTTTGATAATTGAATTAACATTT	63
Dd	3171	AAGTAGGGGTGCAGTGTATTTGTTTTTTTTTATTTTGTGTTTTATTTAATTAAG	3230
OY	64	TTTTTGGCCCAATGATATATAAAAATTTGGATAAATAATATATGATATTCGTTTTTA	123
Dd	3231	TTTGAATAGTTATTTATTTATAGAGTTTTTGTTGGAGTTTTTTTGGAAATTAATAATTAAT	3290
OY	124	GTTGGGTTTGAGAAAAAGSGTTTGACCTTTGGAAAGTGCAGCATGTATATAGATTTGGAG	183
Dd	3291	TATGAGATTTATTAAGTTTAGTTTATGGAAGAAGAAATTTGGAAAGGAATTTTGGGTG	3350
OY	184	CTAGGTGAGCTTTGGACATTTGATTTGATGTGTTGATTTATTAAGTGCACACATTT	243
Dd	3351	GGATTTTAGATATTTATTTATTTTGTGTTGATTTGATGATTAATAATTTTAGTTTAAT	3410
OY	244	AAACCTTAAATGGCTTTCTATAGAGGCCCATTAATTTAGCATTTATACAAGTGACMAC	303
Dd	3411	TTAAACGAAGTGGCGTTGGTGGAGGATTTGGGTAAGTAGCGTTGGAAGATATTTAGTAT	3470
OY	304	TTTTACTTCCTTTTTTTCGACGACATTA	333
Dd	3471	TATATTTGATTTTGGATGATTTTAAATAA	3500

RESULT 6
US-10-311-455-1

; sequence 1, application US/10311455
; GENERAL INFORMATION:

```

1 APPLICANT: OLEK, Alexander
2 APPLICANT: PIEPENBROCK, Christian
3 APPLICANT: BERLIN, Kurt
4 TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De
5 TITLE OF INVENTION: cytosine methylation
6 FILE REFERENCE: 5013.1014
7 CURRENT APPLICATION NUMBER: US/10/311,455
8 PRIOR FILING DATE: 2002-12-16
9 PRIOR APPLICATION NUMBER: PCT/EP01/07537
10 PRIOR FILING DATE: 2001-07-02
11 PRIOR APPLICATION NUMBER: DE 10032529.7
12 PRIOR FILING DATE: 2000-06-30
13 PRIOR APPLICATION NUMBER: DE 10043826.1
14 PRIOR FILING DATE: 2000-09-01
15 NUMBER OF SEQ ID NOS: 2424
16 SEQ ID NO 1
17 LENGTH: 7351
18 TYPE: DNA
19 ORGANISM: Artificial Sequence
20 FEATURE:
21 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
22 US-10-311,453-1

```

Query Match	15.4%	Score 55;	DB 9;	Length 7351;
Best Local Similarity	48.0%;	Pred. No. 0.61;		
Matches 157;	Conservative 0;	Mismatches 170;		

OY	20	TATTTATATCTCTCTCTTTTTTTTTTGATAATTTTAAATCATTTTTTTTGGCCAAATGATA	79
Db	3765	TTTCTATTTCTATTTGGTATAGTTATGGAAAGTTTTTTTTTATTTTGGTTTTTCGAGATA	382
OY	80	TATATAATTTGGATATAATATATATTTGATATTCGTTTTTTTGTTCGGGCTTTGGAAA	139
Db	3825	ATTATATTTATATATTTTATAGATTATTTATATATTTGTTTTTTTTTAAAAATTTTTTAAA	388
OY	140	AGGGTTTCGACTTTCGAAAGTGGACAGTCATATATAGATTTGGGACCTAGTTGAGCTTTG	199
Db	3885	AGGTTTTAGGTTTCGGTTATCTGTTTTTTTATNGCTTTTATAGATATTTTGAAATTTTTTCT	3944
OY	200	GACATTTGATTTGGATGTTGTTGATTTATTAAGTGTGACACTATTTAAACCTTAATAGGGCT	259
Db	3945	AATATAATGTTGGTATTTGATTTATATATAGTATTTATATTTGATTTATTTATATGAT	4004
OY	260	TTCCTAATAAGGCCAATTTATTTAGCATTTAACAACCAAGTGCACACTTTTACGCTTTTGG	319
Db	4005	TTTATATAGTAGGTATTTATTTGATTTTATATATAGTAGTGTGTTTTTTGTTGGTTTTGCT	4064
OY	320	ATCCGACGCAATTAACAATTTGTCAAAT	346
Db	4065	TTTAAATAGAAATGTAATAATTTTTTAGCT	4091

RESULT 7
US-10-24

; Sequence 164, Application US/10240485
; GENERAL INFORMATION.

1 APPLICANT: OLEK, Alexander
 2 APPLICANT: PIEPENBROCK, Christian
 3 APPLICANT: BERLIN, Kurt
 4 TITLE OF INVENTION: Diagnosis of Diseases Associated with
 5 TITLE OF INVENTION: Metastasis
 6 FILE REFERENCE: 5013.1007
 7 CURRENT APPLICATION NUMBER: US/10/240,485
 8 CURRENT FILING DATE: 2002-10-02
 9 PRIOR APPLICATION NUMBER: PCT/EP01/03970
 10 PRIOR FILING DATE: 2001-04-06
 11 PRIOR APPLICATION NUMBER: DE 10019058.8
 12 PRIOR FILING DATE: 2000-04-06
 13 PRIOR APPLICATION NUMBER: DE 10019173.8
 14 PRIOR FILING DATE: 2000-04-07
 15 PRIOR APPLICATION NUMBER: DE 10032529.7
 16 PRIOR FILING DATE: 2000-06-30
 17 PRIOR APPLICATION NUMBER: DE 10043826.1

```

? PRIOR FILING DATE: 2000-09-01
? NUMBER OF SEQ ID NOS: 202
? SEQ ID NO 164
? LENGTH: 20486
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-164

Query Match      15.4%; Score 55; DB 9; Length 20486;
Best Local Similarity 49.6%; Pred. No. 0.52;
Matches 169; Conservative 0; Mismatches 170; Indels 2; Gaps 1;

QY 17 AACGATTTTATCTCTCTCTTTTGGATTAATTTAAATCATTTTGGCCCAATG 76
DB 11836 ACGTAAATGATGATTTTAAATTAATTTAGTAAATATGTTTAAAGTTTTTTTAAAAA 11895
QY 77 AATATATAAATTTGGATTAATATATATTTGATATCGTTTGTAGTCCGGTTGAG 136
DB 11896 GTAAAGAAAAATTTAGATTAATATATATTTAGATTTTAACTTTTAAATGCTTTT 11955
QY 137 AAAAGGCTTTCGACTTTCGAAAGTGACGATGTATATAGATTGGAGCTAGGTGACT 196
DB 11956 TTGTTTTCGAATTTGCTAAAGTTTAAATATTAATTTGATGTTTTTTATTTT 12015
QY 197 TTGGACATTTGTTTGGATGTGTTGATTAATTTAGTGTGACACTATTAACCTTAATG 256
DB 12016 TTAGAGGTTTAAATAGTATATTTAGTTAGTAGAGATGACATTTAAATTAATA -TAAATTT 12073
QY 257 GCTTGTATTAAGGCCCAATTAATATAGCATTAATACAAAGTGACAACTTTACTGCTT 316
DB 12074 GATATAAATGATTTATTCGTAATAGTATTAATTAATTAAGATTAATAATTTTAA 12133
QY 317 TTGATCCGAAGCAATTAACAATTTGCAATTTACCAACACAA 357
DB 12134 TTATGTTAAAAAATATTAATTTATTTTAATTTGAAATTTAA 12174

RESULT 8
US-10-312-841-2
? Sequence 2, Application US/10312841
? GENERAL INFORMATION:
? APPLICANT: Epigenomics AG
? TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
? FILE REFERENCE: E01/1208/MO
? CURRENT APPLICATION NUMBER: US/10/312, 841
? PRIOR FILING DATE: 2002-12-30
? NUMBER OF SEQ ID NOS: 2
? SEQ ID NO 2
? LENGTH: 3673778
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (379615)
US-10-312-841-2

Query Match      15.3%; Score 54.8; DB 9; Length 3673778;
Best Local Similarity 47.6%; Pred. No. 0.26;
Matches 161; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 8 AAAAGATGAAGATTTTATCTCTCTCTTTTGGATTAATTTTAAATCATTTT 67
DB 2157283 AATGAAAAATTTTGAGTTATTTTTCGTATGTTTATTTTGTAGCGTGTCTTAT 2157342
QY 68 TGCCCAATGATATATATAAATTTGGATTAATATATTTATGATATTCGTTTGTAGTTC 127
DB 2157343 TATTTATATTTTAAAGTGAAGTGTGCTGTAGAGATATGCTGTTGTTGCTGTG 2157402
QY 128 GGGTTTGACAAAAAGGCTTTCGACTTTCGAAAGTGACGATGTATATAGATTGGAGCTAG 187

```

```

Db      2157403 TTTTCTAGTGATGGGGAATTAAGGTGGAGAGGGTGAACCAACGGTTTTTCGGGAGACAGA 2157462
Qy      188 GTTAGCTCTTTGGACATTGTAATGAGATGTTGATTAATTAAGTGCACACTAATTAAC 247
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2157463 GTAGAGAAATAATGTAATTAATTTATTTGTTAATAGTCTTTATATATTTTTCGGGAATTAATTA 2157522
Qy      248 CTAAATAGCGCTTCTATTAAGGCCCAATATATATACGATTTATACAAAGTGCACACTTT 307
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2157523 TAATTTAAGTGTGTTTAATTTAAAGGTAAGTGTGTTTAAAAAAAATAATATGTAATTT 2157582
Qy      308 ACTGCTTTTGTGATCCGAAGCATATAACAATATGTCAAA 345
        ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2157583 ATTTATTTGAGGTAGAAATTTAAAAAATATATATATA 2157620

RESULT 9
US-10-311-455-2100
; Sequence 2100, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2100
; LENGTH: 2814
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2100

Query Match          15.1%; Score 54.2; DB 9; Length 2814;
Best Local Similarity 48.8%; Pred. No. 0.97;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy      20 TATTTTATTCCTCCTTTTGTGTAATTTTAANCATTTTTTGGCCANAGTA 79
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3 TTTGTTTAATTTAATTTAATTTAATTAATATATATATATATATATATATATATATATATATATAT 62
Qy      80 TATAAAATTTGGATAATTAATATATATGATATCGTTTTTACTGCGGTGAGAAA 139
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      63 TAGCGTAGGCGGAGATATATATTTTGTATATATTTTCGTATATAGTAATTTTGTAT 122
Qy      140 AGCGTTTCACCTTCGAAAGTGACAGATATATAGTTGGAGCTAGCTGCTTGG 199
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      123 ATGTAATATGTAATATATATATATATATATGTAATGTAATATAGCTTTTGTATAAACGTTATA 182
Qy      200 GACATTTGTATGATGTTGTGATATTAAGTCGACATTTAAACCTTAATGGCGT 259
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      183 TGTAATATATATATATATATATTTTATATATAGTTTATATATATATATATATATATATATATATAT 242
Qy      260 TTCTATAGGCCCAATTTATTTACGATTTAAACAAGTACAACTTTACTGCTTTT 318
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      243 TTTGGGAAGTTTATATATTTTATAGATTATATAAGGAGATTTATTTTATTTT 301

```

```

? PRIOR APPLICATION NUMBER: DE 10043826.1
? PRIOR FILING DATE: 2000-09-01
? NUMBER OF SEQ ID NOS: 132
? SEQ ID NO 39
? LENGTH: 14147
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
? US-10-311-506-39

Query Match          14.9%; Score 53.2; DB 9; Length 14147;
Best Local Similarity 48.1%; Pred. No. 1.1; Mismatches 0; Gaps 0;
Matches 151; Conservative 0; Indels 163;

QY      7  AAAAAGATCAAACTATTATTATCTCTCTCTTTTGGATAATTTTAAATCAATTTT 66
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10811 AAAAAGTTGGTTAGTTGGTTTGGTTTGGTTATTTGGTTTTTTTTCGTTGTTTTTTT 10870

QY      67  TTGCCCAATGATATATATAAAATTGGATAAATAATATTAATGGAATTCGTTTTAGTT 126
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      10871 TTTTTCGAATTAATGTTAGTTTCTTTTCTTTTATGTAATGTTAGGTAATTTTATGTT 10930

QY      127  CGGGTTTGAGAAAAAGGTTTCGACTTTCGAAACTGGACATGATATTAATTTGGAGACTA 186
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      10931 TTGGGTTTTTTAATGCGTTGTCGTTTATTTATTTGGTTTGGTTTGAATATTTACGAG 10990

```

Db 10991 TTTAGTCTTTTATTTTCTTTGCTTTCTTACGTTGTTTTTTTGGCCGATATTTT 11050

```

Db      11051 TTTTTCATTTTATGTGAAGAAGATAGTTTTTTTGTTCATTTTTTTTATATATTT 11110
QY      307 TACTTCGTTTTGA 320
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      11111 TTATTATTATTGA 11124

RESULT 12
US-10-311-507-39
; Sequence 39, Application US/10311507
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and nucleic acids for the analysis of astrocytomas
; FILE REFERENCE: 5013.1013
; CURRENT APPLICATION NUMBER: US/10/311,507
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07538
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30

```

```

: NUMBER OF SEQ ID NOS: 136
: SEQ ID NO 39
: LENGTH: 14147
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-507-39

Query Match          14.9%; Score 53.2; DB 9; Length 14147;
Best Local Similarity 48.1%; Pred. No. 1.1;
Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY      7  AAAAAGATGAAAGATTTTATCTCTCTCTTTTTTTTCGATTAATTTAAATCATTTT 66
        |||||  |  |||  |  |||  |  |||  |  |||  |  |||  |  |||||
db      10811 AAAAAGTTGGTTAGTTGGTTTGCATTAATTTGGTTTTTTTTTTCGTTGGTTTTTTT 10870

```

[illegible]

; LENGTH: 5678

TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1111

Query Match 14.7%; Score 52.8; DB 9; Length 5678;
Best Local Similarity 54.0%; Pred. No. 1.5;
Matches 108; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 36 TTTTGTGATTAATTTTAAATCATTTTGGCCCAATGATATATAAATTGGATA 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3308 TTTTATTTTATTTTAAATAAGTTTGTATTTTATTTTATTTTATTTTGG 3367
QY 96 AATTAATTAATGATATTCGTTTTTGTGAGTTGAGAAAGGTTTCGACTTCG 155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3368 AATTTTAAATGATATTTTAAATGATGTTTAAAGTTTGTGATTTTGG 3427
QY 156 AAAGTGACGATGATATGATGAGCTAGCTTGTGACATTGTGATGAT 215
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3428 TAAATTTTAAATTTTATTTTATTTGATTTGGGTAATTTTAAATGATTTGATTTGAG 3487
QY 216 GTGTTGATTAATGATGTCG 235
||||| ||||| ||| |||
Db 3488 TATGTTGATTTTATTTTGG 3507

Search completed: June 11, 2003, 07:57:08
Job time : 271.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 542.722 Seconds

(without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcataaccacacaaag 358

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estimu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlum:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_dln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_todl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137.4	38.4	176	17	AL762543 Arabidops
2	128.4	35.9	194	17	AL762544 Arabidops
3	121	33.8	423	10	AV787618 Arabidops
4	68	19.0	1101	17	AL078714 Drosophill
5	65.4	18.3	928	17	CNS00DKY
6	60	16.8	905	17	CNS00KHX
					AL077798 Drosophill

C	7	58.2	16.3	1027	17	CNS02R50	AL127233 Tetradon
C	8	58	16.2	1101	17	CNS016L1	AL106896 Drosophill
C	9	57.8	16.1	961	17	A2691838	AL2691838 ENTM156TF
C	10	57	15.9	1101	17	CNS00DEJ4	AL069257 Drosophill
C	11	56.6	15.8	894	17	CNS018BG	AL1091926 Drosophill
C	12	56.6	15.8	1147	17	CNS073CX	AL4273351 clone BA0
C	13	56.4	15.8	1101	17	CNS0029N	AL097397 Drosophill
C	14	56.2	15.7	1101	17	CNS0021J	AL061936 Tetradon
C	15	56	15.6	949	17	CNS04A1H	AL1755924 Drosophill
C	16	55.8	15.6	767	17	CNS00A0X	AL1053526 Tetradon
C	17	55.6	15.5	1092	17	CNS020K7	AL175596 Tetradon
C	18	55	15.4	1007	17	CNS06X9S	AL19462 T3 end of
C	19	54.8	15.3	1101	17	CNS001FB	AL060732 Drosophill
C	20	54.6	15.3	952	17	CNS006V8	AL065867 Drosophill
C	21	54.6	15.3	1002	17	CNS0185X	AL108927 Drosophill
C	22	54.6	15.3	1121	17	CNS01623	AL106197 Drosophill
C	23	54.4	15.2	423	17	CNS001W9	AL075520 Drosophill
C	24	54.2	15.1	866	17	CNS01161	AL099876 Drosophill
C	25	54.2	15.1	1101	17	CNS00EVL	AL069706 Drosophill
C	26	54	15.1	952	17	CNS016RS	AL107122 Drosophill
C	27	54	15.1	993	9	AL561361	AL561361 AL561361
C	28	54	15.1	1200	17	CNS016CO	AL106578 Drosophill
C	29	53.6	15.0	1043	17	CNS0145P	AL103735 Drosophill
C	30	53.6	15.0	1094	17	CNS012F2	AL101513 Drosophill
C	31	53.6	15.0	1167	17	CNS07360	AL427102 clone BA0
C	32	53.4	14.9	973	17	CNS071LE	AL425064 clone BA0
C	33	53.4	14.9	974	17	AC043196	AC043196 Pan troy1
C	34	53.2	14.9	458	9	AL514085	AL514085 AL514085
C	35	53.2	14.9	994	17	CNS0058N	AL057169 Drosophill
C	36	53	14.8	573	17	CNS06LTX	AL404635 T7 end of
C	37	53	14.8	813	17	CNS03CIW	AL237857 Tetradon
C	38	52.8	14.7	637	17	CNS036CC	AL229845 Tetradon
C	39	52.6	14.7	1225	17	CNS0161D	AL106171 Drosophill
C	40	52.4	14.6	529	9	AL514657	AL514657 AL514657
C	41	52.4	14.6	1101	17	CNS0177R	AL107697 Drosophill
C	42	52.4	14.6	1201	17	CNS01038	AL098462 Drosophill
C	43	52.2	14.6	403	17	AC080113	AC080113 HS 4868_B
C	44	52.2	14.6	759	17	CNS060XV	AL411257 T7 end of
C	45	52.2	14.6	765	17	BH148938	BH148938 ENTP52TF

ALIGNMENTS

RESULT 1	AL762543	176 bp	DNA	linear	GSS 19-JUN-2002
LOCUS	AL762543				
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-025G03-013722,				
ACCESSION	AL762543				
VERSION	AL762543.1	GI:21508636			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
TITLE	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
1	Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.				
2	and Weissshaar, B.				
3	A pipeline for automated high-throughput generation of FSTs				
4	(flanking sequence tags) from Arabidopsis thaliana T-DNA				
5	transformed lines				
6	Unpublished				
JOURNAL	Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.				
REFERENCE	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)				
AUTHORS	for flanking sequence tag based reverse genetics				
TITLE	Unpublished				
JOURNAL	3 (bases 1 to 176)				
REFERENCE	Li, Y., Strizhov, N., Rosso, M. and Weissshaar, B.				
AUTHORS	Direct Submission				
TITLE	Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer				
JOURNAL					

COMMENT
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone 13A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
Location/Qualifiers
1. 176
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025603-013722"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT
ORIGIN
60 a 33 c 27 g 56 t

Query Match
Best Local Similarity 38.4%; Score 137.4; DB 17; Length 176;
Pred. No. 1e-11; Mismatches 0; Indels 1; Gaps 0;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 220 TTGATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATTATA 279
1 TTTATTTATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATTATA 60

Db 280 TTACGATTATTAACAAGTGAACCTTTACTCTGTTTTCGCGAAGCATATTAACATT 339
1 TTTATTTATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATTATA 120

QY 61 TTACGATTATTAACAAGTGAACCTTTACTCTGTTTTCGCGAAGCATATTAACATT 120

QY 340 GTCAATATACCAACACAG 358
121 GTCAATATACCAACACAG 139

Db 121 GTCAATATACCAACACAG 139

RESULT 2
AL762544 194 bp DNA linear GSS 19-JUN-2002
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-025603-013757,
genomic survey sequence.
ACCESSION AL762544
VERSION AL762544.1 GI:21508641
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Strizhov,N., Li,Y., Rosso,M., Vlehever,P., Dekker,K., Saedler,H.
and Weishaar,B.
TITLE A pipeline for automated high-throughput generation of ESTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weishaar,B.
TITLE A new Arabidopsis thaliana T-DNA mutagenesis population (GABI-Kat)
for flanking sequence tag based genetics
JOURNAL Unpublished
REFERENCE
AUTHORS Li,Y., Strizhov,N., Rosso,M. and Weishaar,B.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weishaar B., Max-Planck-Institut fuer

COMMENT
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone 13A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
Location/Qualifiers
1. 194
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025603-013757"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT
ORIGIN
66 a 37 c 28 g 63 t

Query Match
Best Local Similarity 35.9%; Score 128.4; DB 17; Length 194;
Pred. No. 2.1e-10; Mismatches 1; Indels 1; Gaps 1;
Matches 140; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 217 TTGTTGATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATT 276
1 TTTATTTATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATT 85

Db 27 TTGTTGATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATT 85

QY 277 ATATTAGATTATTAACAAGTGAACCTTTACTCTGTTTTCGCGAAGCATATTAACATT 336
1 TTTATTTATTTAGTGTGCGACATATTAAACCTTAATGGCTTTCTATAGGCCCAATT 145

Db 86 ATATTAGATTATTAACAAGTGAACCTTTACTCTGTTTTCGCGAAGCATATTAACATT 145

QY 337 ATTGTCAATATACCAACACAG 358
146 ATTGTCAATATACCAACACAG 167

Db 146 ATTGTCAATATACCAACACAG 167

RESULT 3
AV787618 423 bp mRNA linear EST 28-MAR-2002
LOCUS AV787618 RFLP6 Arabidopsis thaliana cDNA clone RFLP6-76-F19 3',
mRNA sequence.
ACCESSION AV787618
VERSION AV787618.1 GI:19806408
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 423)
REFERENCE
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carlini,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shingawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rcc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). This clone is in a

modified pBluescript vector as a SstI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/Plant/index_e.html) for further details.

FEATURES

SOURCE

Location/Qualifiers
1. 423

/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAF106-76-F19"
/clone_lib="RAF16"
/dex_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: SstI; site 2: XhoI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT 156 a 79 c 64 g 124 t
ORIGIN

Query Match 33.8%; Score 121; DB 10; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGATGAAAGTATTTTATCTCTTTTGTGATATTTAAATCA 60
|||
Db 121 AAAAAAAAAAGATGAAAGTATTTTATCTCTTTTGTGATATTTAAATCA 62-
61 TTTTGTGCGCATGATATATAAATTTGATATAATATATGATATTCGTTT 120
|||
OY 121 T 121
||
Db 1 T 1

RESULT 4
CNS00LT2 1101 bp DNA linear GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR4BP19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL078714.1 GI:5102004
VERSION AL078714
KEYWORDS GSS.

SOURCE

Drosophila melanogaster.

ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.

REFERENCE
AUTHORS Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE BP 191 91006 EVRI cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
JOURNAL - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE

Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"
/clone="BACR4BP19"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN
Query Match 19.0%; Score 68; DB 17; Length 1101;
Best Local Similarity 26.4%; Pred. No. 0.1;
Matches 68; Conservative 100; Mismatches 90; Indels 0; Gaps 0;

OY 4 AAAAAAAAAAGATGAAAGTATTTTATCTCTTTTGTGATATTTAAATCA 63
|||
Db 603 WAKWMAWMDATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 662
64 TTTTGTGCGCATGATATATAAATTTGATATAATATATGATATTCGTTT 123
|||
Db 663 TTTTGTGCGCATGATATATAAATTTGATATAATATATGATATTCGTTT 124
|||
OY 124 GTTGGGTTTGAGAAAGCGTTTGACTTTGAGAAAGCGAGATGATATGATGGAG 183
|||
Db 723 KATKTKKKKKAAWAAADKKRRKGGKKKKKKKKKKKKKKKKKKKKKK 782
|||
OY 184 CTAGTTGAGCTTTTGACATTTGATTTGATTTGATTTAGTGACATAT 243
|||
Db 783 AGDDAKDKTKKKKKKATTTTCKKKKKKKKKKKKKKKKKKKKKKK 842
|||
OY 244 AAACCTTAATGCGCTT 261
|||
Db 843 TKDKKKKKKKTKTKTK 860

RESULT 5
CNS00DKY 928 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865
KEYWORDS GSS.

SOURCE

Drosophila melanogaster.

ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 928)
Genoscope.

REFERENCE
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE BP 191 91006 EVRI cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
JOURNAL - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE

Location/Qualifiers
1. 928
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCI-98"

BASE COUNT	518 a	63 c	91 g	289 t
ORIGIN				
Query Match		16.1%	Score 57.8;	DB 17; Length 961
Best Local Similarity		64.7%;	Pred. NO. 3.8;	

	Matches	86;	Conservative	0;	Mismatches	47;	Indels	0;	Gaps	0;
OY	1	AAAAAAAAAAGATGAAGAATTCTTCGTTTGTGGATTAATTTAAATCA	60							
Dd	674	AAAAAAAAAAAAATTTTATTTTATTTTATTTTATATAATTTAAAACA	615							
OY	61	TTTTTTTTGCCCATGATATATAAATTTGCATTAATATATTCGATTCTT	120							
Dd	614	TTTTTATATTTTATTTATATATAAATATTTAAAAATTTTTTTTTTTTT	555							
OY	121	TTAGTGCGGCTT	133							
Dd	554	TTTTTTTTTTTTT	542							
RESULT 10 CONS0DEJ4/C LOCUS	CNS0DEJ4	1101 bp DNA linear GSS 04-JUN-1999								
DEFINITION	CNS0DEJ4 melanogaster genome survey sequence T7 end of BAC:									
ACCESSION	BACR2K22 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.									
VERSION	AL069257									
KEYWORDS	AL069257.1 GI:4949400 GSS.									
SOURCE	Drosophila melanogaster. Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidae; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)									
REFERENCE	determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see "http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammocci at Pleter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the EscoRI strain YZ; cn bw sp, the same strain used for the BDGP's pI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.									
AUTHORS	Location/Qualifiers									
TITLE	1. 1101									
JOURNAL	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR2K22" /clone_lib="RPCI-98" /note="end : T7"									
COMMENT										
FEATURES	source									
BASE COUNT	434 a 140 c 25 g 377 t 125 others									
ORIGIN										
Query Match	15.9% Score 57; DB 17; Length 1101;									
Best Local Similarity	42.1%; Pred. No. 4.6;									
Matches	149; Conservative 35; Mismatches 163; Indels 7; Gaps 1;									
OY	1	AAAAAAAAAAGATGAAGAATTCTTCGTTTGTGGATTAATTTAAATCA	60							
Dd	895	AAAMAAATRAATTTTTTAAAAATTTTWTWATTTTTTTTAAAAAATTTWAATTT	836							
OY	61	TTTTTTTGGCCATGATATATAAATTTGCATTAATATATTCGATTCTT	120							
Dd	835	TWATTTTAAADADAANAAMDNAANAGCAAAAAAGCAATTTTTCGTAANAAAAAWA	776							
OY	121	TTACTTCGGGCTT-----GAGAAAAGGGTTTCGACTTGGAAAAGTGACGATGTAT	173							

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES		
Db	CNS018BG/c	Drosophila melanogaster genome survey sequence T7 end of BAC BACN13B16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	AL109126	GI:5629430	GSS.	Drosophila melanogaster.	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	1 (bases 1 to 894)	Genoscope.	Submitted Submission	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)	Location/Qualifiers		
Db	775	TTWTAATGAGTGGTGGGAGMTAAGGGGTWATTTTATATACATTTATTTAATTTT	174	AGATGGGAGCTAGCTGAGTCTTTGGACATTTGATTTGGATTTGATTTAGTGT	715	TTWTTTGAAGAAGTAATGTTGAATATTTTATKATGGTATATGTTTTTTTTTATATAT	234	CGACATATTAACCTTAATATGCGCTTTCATATAGCCCAATTAATATTAACATTA	655	TKTTTAGAAAAATATTAATTAATGTTTTTTTGTATAGAAAGAAATTAATTAATTA	294	AAGTGACAACTTACTCGTTTTTGATCCGAGCAATTAACAAATTTGCAATA	595	ATGNNNTGCGTCTTTTGGTTTTTATTAATTAATTAATTAATTAATTAATTAATTA
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;	DB 17;	Length 894;										
Best local Similarity	38.4%;	Pred. No. 6;												
Matches 101;	Conservative 44;	Mismatches 118;	Indels 0;	Gaps 0;										
Db	284	TT	80	TATTAATTTGCTCTCTCTTTTGTGTAATTTTAATTCATTTTTTTTGGCCATGATA	224	TTTWTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	140	AGGCTTTGAGCTTTGGAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	164	ACCTAAATGCTTTGCAAGTGAAGATGATATATAGATTGGAGCTAGCTGAGTCT	200	GACATTTGATGATGTTGTTGATATATATAGTGCAGACATTAATTAACCTTAAT		
Qy	15.8%;	Score 56.6;</												

Db	685	AANTTTAAATTTTTTNAATNTTTTTTTTTTTTTTTGTAATWMAAAAAAATWTTTT 744
QY	70	CCCAAGATATAATAAATTGGATTAATATATATTCATTCGTCTTTTAGTTCG 129
Dd	745	TTTTTTTTTTTTTTATATWMAAAAAATWMAAACAATAATTTTTTKTGGR 804
QY	130	GTTTGAGAAAAGCGTTTGACTTCGAAAGTGACGATGATATAGATGAGGT 189
Dd	805	KATMAAAAMWMAAGWAKAGTDWMAAAAAATAGRTWWMTTATATATWRRTDGGKT 864
QY	190	TGAGCTTTTGACATTTGTATTTGATGTGTGATTTATGTGCGA 236
Dd	865	TTTTTTTTTTTTTAKTTGGGGGKRITTTTTTKKKKKGCKRRKWGA 911

RESULT 13	
CNS0029N	
LOCUS	
DEFINITION	CNS0029N 1101 bp DNA linear GSS 26-JUL-1999
ACCESSION	Drosophila melanogaster genome survey sequence T7 end of BAC
VERSION	BACN01C10 of DrosBAC library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
SOURCE	AL097397
ORGANISM	GI:5609008
	GSS.
	Drosophila melanogaster.
	Drosophila melanogaster.
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryoeta;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 1101)
	Genoscope.
	Direct Submission
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :
	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
	- Web : www.genoscope.cns.fr)
	determination of this BAC-end sequence was carried out as part of a
	collaboration with the European Drosophila Genome Project (EDGP) -
	http://www.edgp.ebi.ac.uk -. This drosophila melanogaster BAC
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre
	d'Etude du Polymorphisme Humain) with funding provided by a MRC
	project grant. The DNA was prepared from embryos by Alain Bucheton
	and Genevieve Payan. It has been constructed in the vector
	pBelobAC11.

FEATURES	
source	Location/Qualifiers
	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_id="BACN01C10"
	/clone_lib="DrosBAC"
	/plasmid="pbelobAC11"
	/note="end : 17"

BASE COUNT	254 a	288 c	267 g	177 t	115 others
ORIGIN					

Query Match	15.8%; Score 56.4; DB 17; Length 1101;
Best Local Similarity	47.4%; Pred No.5.7;
Matches	83; Conservative 24; Mismatches 68; Indels 0; Gaps 0;

QY	2	AAAAAAAAAGAAGATATTTATCTCTCTCTTTTGTATATTTAATCAT 61
Dd	892	WAHA 951
QY	62	TTTTTGGCCAATGATATATAAATTTGGATAAATAATATATGATATTCGTTTT 121
Dd	952	TTTTTGATATWMTTTRKNDAGTAAMAATTTTWTATATTTTTTKRTTTTTTTTTTT 1011
QY	122	TACTGGGGTTTACGAAAGGCTTCGACTTGCAGAAAGTGACGATGATATAGA 176
Dd	1012	TTTTTTTTTATWATGAARTTTTTTTTTTTTAKAAAAAGAWTADAAATTTAA 1066

RESULT 14	
CNS0021J/c	

LOCUS CNS0021J 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR05N11 of Rpct-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL061936
VERSION AL061936
KEYWORDS GI:4940214
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BASE COUNT	631 a	7 c	28 g	289 t	146 others
ORIGIN					

[illegible]

LOCUS	CNS04AIH/C	949 bp	DNA	linear	GSS 21-MAY-2000
DEFINITION	Tetradon nigriviridis genome survey sequence T7 end of clone 095f19 of library G from Tetradon nigriviridis, genomic survey sequence.				
ACCESSION	AL281906				
VERSION	AL281906.1	GI:8020236			
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Tetradon nigriviridis.				
ORGANISM	Tetradon nigriviridis.				

```

BASE COUNT      385 a      144 c      185 g      223 t      42 others
ORIGIN
Query Match          15.6%, Score 56; DB 17; Length 949;
Best Local Similarity 49.3%; Pred. No. 7.1;
Matches 74; Conservative 21; Mismatches 55; Indels 0; Gaps 0;

QY    4   AAAAAAAAAAGATGATATTTTAACTCCCTTTTTTTTTTGATAATTAAACATTT 63
       . : : : : : : : : | | | | | | | | | | | | | | | | | | |
Db     820 WMAAAMAAAAMWMAAMTTTTTTTTTTTTTTTTTTTTWWNNNAATTTTTTTT 761

QY    64 TTTTGGCAATGATATATAAAAAATTGGATAAATATATTTGATATTCGTTTTA 123
       ||||| : : : | : | : | | | | | | | | | | | | | | | |
Db     760 TTTTTTTTTTWMTTWMMAAAAMTTTTTTTTTTTTWWWMAAMTTTAAATTTAAATTTT 701

QY    124 GTTCGGCTTGAGAAAAGCGTTGCACITT 153
        || | | | | | | | : | | |
Db     700 TTTTTTTTTTTTAAAAAMTTTTTTAWMTTT 671

Search completed: June 11, 2003, 06:58:40
Job time : 545.722 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

June 10, 2003, 21:51:23 ; Search time 44 Seconds
(without alignments)
1553.581 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681

Sequence: 1 MFETHTLPLLLPSLTS.....FAFPVDFPGLPIRVSRL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	21	AA07921
2	978.5	36.5	472	18	AAW27153
3	978.5	36.5	472	21	AA044571
4	978.5	36.5	472	21	AA045022
5	978.5	36.5	491	21	AA045021
6	978.5	36.5	492	21	AA044570
7	946	35.3	444	21	AA044572
8	946	35.3	444	21	AA045023
9	866	32.3	512	23	AB091571
10	816.5	30.3	486	22	AAU02839

11	726	27.1	461	21	AA046490	Arabidopsis thalia
12	726	27.1	462	21	AA046489	Arabidopsis thalia
13	726	27.1	465	21	AA046491	Arabidopsis thalia
14	712	26.6	461	21	AA041836	Arabidopsis thalia
15	712	26.6	462	21	AA041835	Arabidopsis thalia
16	712	26.6	465	21	AA041834	Arabidopsis thalia
17	653.5	24.4	468	21	AA020783	Arabidopsis thalia
18	606	22.6	471	21	AA023013	Arabidopsis thalia
19	606	22.6	479	21	AA023012	Arabidopsis thalia
20	603	22.5	469	21	AA023014	Arabidopsis thalia
21	598.5	22.3	485	21	AA020784	Arabidopsis thalia
22	597.5	22.3	485	21	AA020785	Arabidopsis thalia
23	596.5	22.2	497	22	AAU02839	Taxus cuspidata ox
24	592	22.1	496	22	AAU02837	Taxus cuspidata ox
25	588	21.9	509	22	AAU02832	Taxus cuspidata ox
26	585	21.8	507	22	AAU02830	Taxus cuspidata ox
27	583	21.7	484	22	AAU02822	Taxus cuspidata ox
28	571	21.3	501	22	AAU02823	Taxus cuspidata ox
29	569.5	21.2	493	22	AAU02836	Taxus cuspidata ox
30	568.5	21.2	500	22	AAU02824	Taxus cuspidata ox
31	564	21.0	494	22	AAU02821	Taxus cuspidata ox
32	564	21.0	500	22	AAU02826	Taxus cuspidata ox
33	558.5	20.8	483	22	AAU02827	Taxus cuspidata ox
34	558.5	20.8	482	22	AAU02834	Taxus cuspidata ox
35	553.5	20.6	488	22	AAU02838	Taxus cuspidata ox
36	547.5	20.4	489	21	AA030048	Arabidopsis thalia
37	546	20.4	481	21	AA030049	Arabidopsis thalia
38	534.5	19.9	503	21	AA020785	Arabidopsis thalia
39	528.5	19.7	388	21	AA020785	Arabidopsis thalia
40	503	18.8	492	19	AA037733	Arabidopsis thalia
41	503	18.8	492	19	AA041459	Cyrtocarpus z450RA
42	503	18.8	492	22	AA085155	Zebrafish cytochrome
43	503	18.8	492	23	AA085155	Zebrafish cytochrome
44	490	18.3	433	21	AA030050	Arabidopsis thalia
45	461	17.2	497	19	AA037735	Cyrtocarpus z450RA

ALIGNMENTS

```

RESULT 1
ID      AAB07921 standard; Protein; 513 AA.
XX
AC      AAB07921;
XX
DT      14-NOV-2000 (first entry)
XX
DE      A cytochrome P450 enzyme designated DMF4.
XX
KW      DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
XX      plant phenotype; cell elongation.
XX
OS      Arabidopsis sp.
XX
PN      WO200047715-A2.
XX
PD      17-AUG-2000.
XX
PF      11-FEB-2000; 2000WO-US03820.
XX
PR      11-FEB-1999; 99US-0119657.
XX
PT      11-FEB-1999; 99US-0119658.
XX
PI      (ARIZ-) ARIZONA BOARD OF REGENTS.
XX
PI      Azpiroz R, Choe S, Feldmann KA;
XX      WPI; 2000-549142/50.
XX      N-PSDB; AAA59599.
XX      New isolated dwf4 polynucleotide useful for altering the phenotype of
XX      plants, for diagnostic assays and in the production of antibodies -
XX

```


DB 447 DKLVFFPTTRQKRPFFVKR 467

RESULT 3
AAG44571
ID AAG44571 standard; Protein: 472 AA.
XX AAG44571;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139730.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175-
PR 18-AUG-1999; 99US-0149426-
PR 20-AUG-1999; 99US-0149722-
PR 20-AUG-1999; 99US-0149723-
PR 20-AUG-1999; 99US-0149929-
PR 23-AUG-1999; 99US-0149902-
PR 23-AUG-1999; 99US-0149930-
PR 25-AUG-1999; 99US-0150566-
PR 26-AUG-1999; 99US-0150884-
PR 27-AUG-1999; 99US-0151065-
PR 27-AUG-1999; 99US-0151066-
PR 30-AUG-1999; 99US-0151080-
PR 31-AUG-1999; 99US-0151303-
PR 01-SEP-1999; 99US-0151930-
PR 07-SEP-1999; 99US-0152363-
PR 10-SEP-1999; 99US-0153070-
PR 13-SEP-1999; 99US-0153758-
PR 15-SEP-1999; 99US-0154018-
PR 16-SEP-1999; 99US-0154039-
PR 20-SEP-1999; 99US-0154779-
PR 22-SEP-1999; 99US-0155139-
PR 23-SEP-1999; 99US-0155486-
PR 24-SEP-1999; 99US-0155659-
PR 28-SEP-1999; 99US-0156458-
PR 29-SEP-1999; 99US-0156596-
PR 04-OCT-1999; 99US-0157117-
PR 05-OCT-1999; 99US-0157753-
PR 06-OCT-1999; 99US-0157865-
PR 07-OCT-1999; 99US-0158029-
PR 08-OCT-1999; 99US-0158232-
PR 12-OCT-1999; 99US-0158359-
PR 13-OCT-1999; 99US-0159293-
PR 13-OCT-1999; 99US-0159294-
PR 13-OCT-1999; 99US-0159295-
PR 14-OCT-1999; 99US-0159329-
PR 14-OCT-1999; 99US-0159330-
PR 14-OCT-1999; 99US-0159331-
PR 14-OCT-1999; 99US-0159637-
PR 14-OCT-1999; 99US-0159638-
PR 18-OCT-1999; 99US-0159584-
PR 21-OCT-1999; 99US-0160741-
PR 21-OCT-1999; 99US-0160767-
PR 21-OCT-1999; 99US-0160768-
PR 21-OCT-1999; 99US-0160770-
PR 21-OCT-1999; 99US-0160814-
PR 21-OCT-1999; 99US-0160815-
PR 22-OCT-1999; 99US-0160980-
PR 22-OCT-1999; 99US-0160981-
PR 22-OCT-1999; 99US-0160989-
PR 25-OCT-1999; 99US-0161404-
PR 25-OCT-1999; 99US-0161405-
PR 26-OCT-1999; 99US-0161406-
PR 26-OCT-1999; 99US-0161359-
PR 26-OCT-1999; 99US-0161360-
PR 26-OCT-1999; 99US-0161361-
PR 28-OCT-1999; 99US-0161920-
PR 28-OCT-1999; 99US-0161992-
PR 28-OCT-1999; 99US-0161993-
PR 29-OCT-1999; 99US-0162142-
Query Match 36.58; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.94; Pred. No. 6; Ge-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLLSLLLKRLRRKTRFLLPPGSGWPLGFTIGYIKRYTATITLGDPMQOH 71
Db 7 LLLSSIAAGFL---LLKRTTRRRMGGLPPGSLGLPLIGTFOLIGAKTNEPPEPTDER 63
QY 72 VSKYKGIYRSNLTGEPITVSADAGLNKRFILONEGRLFECSYPRSIGILGKWSMLVGD 131
Db 64 VARYGSVFWTHLGEPTIISADPETNRFLVONEGKGLFECSYPRASICNLGKHSLLMKGS 123

QY 132 MHRDMRSTSLNPLSHARLRTILKQVERHTLFLVDSMOONSIFSAODEAKKFTENLMK 191
Db 124 LKRMHSITLMSFANSSIIDHILMDIDLRLVFNLDLSWMSRVLL--MEBAKKTIFELTVKQ 181
QY 192 IMSMDPGHEETEOAKKEVTFEMKGVSAPLNPGFAYKALOSRATILKFERKMEERKL 251
Db 182 LMSFDPC-EMESLURKELLYIEGFSLLPLFSTTYKALQAR-----RKVAEALT 232
QY 252 DIKEEDQEEBEVKTEDEAEMSKSDHVRKQRTDDLLGWLKHSNLSFEOILLDLISLFA 311
Db 233 VVYMKRREEEGAE-----RKMDLAALLADDGFSDEIDYFVALLVA 278
QY 312 GHETSSVAIALAIFFLQCPRAVEELREHLEIARAKKELGSELMNDYKKMDTQCVI 371
Db 279 GYETSTTMTLAVKFLTETPLALQKEHEKIRAMKSD--SYLSWSDYKSMPTOCVV 336
QY 372 NETRLGVNVRFLHKKALKDVRKGYDIPSGKVLVPSAYHLONSRYDQPLFPMHMO 431
Db 337 NETLRVANTIGGVRRAMTDVEIKQKIPKGVSSFRVAHLDPNHEKDKARTFPMHMO 396
QY 432 QONNGASSGSGSFTWGN--YMPGCGPRLCAGSELAKLMAVFIHHLVLFKNMELAD 490
Db 397 -----SNSVTGPGSNVFTPGGGRPLCQPELARVALSVFLHRLVTFGFSVPAEQ 446
QY 491 DQPAFPVDFPNGLPIRVSR 511
Db 447 DKLVEFPPTRTQKRPPIVKR 467

RESULT 4

ID AAG45022 standard; Protein; 472 AA.

AC AAG45022;

DE 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 56469.

KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 23-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 12-AUG-1999; 99US-0148171.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 6.6e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLLFLILKRNKRTRFLPGKSGWPLGFTIGLYTKYTTTADENQOH 71
DB 7 LLLLSIAAGFL--LLRTRRMRGLPGSLGLPLIGFTFOLIGAYKTENPEPTIDER 63
QY 72 VSKYKTYRNLGPEPTIVADAGLNRFILQNGRLEEGSPRSIGILGKMSLVVGD 131
DB 64 VARYGVFHTLGEPTIESADPETNRFLVQNEGKLEFECSPASICMLLKRHSLLMKGS 123
QY 132 MHRDMSISINLSHARLRTILKLDVERHTLFYLDWQONSIFSADAKKFTFNMAKH 191
DB 124 LHKRHSITWSPANSSITKIDMLDRLVFNLDSSSHVL--MEAKKTFPELVKQ 181
QY 192 IMSMDGEETEDLKKRYTFMGVVSAPLPGTAHKLQSAITLKTIERMERKL 251
DB 182 LMSFDG-EMSESLRKEVLYIGFSLPLSTYRKAIQAN-----RKVAALP 232
QY 252 DIKEDEEEETEDAEKMSKDHVKQRTDILLGVKHSNLSTEQIIDLILSLFA 311
DB 233 VVAKRREBEAE-----RKQMLALLAADGFSDEIVDFVALLVA 278
QY 312 GHTSSVAIALAIFLQACKAVEELREHLEIARAKKEIGESELNMDYKKDFTQCVT 371
DB 279 GYETSTIMLAKFLETPPLAOLKEHEKIRAMKSD--SYLSMEDSKSMPTQCV 336
QY 372 NETLRGNVRFHRAKALQVRKYGDIPSGKVLPIVSIYVHLDNSYDQNLFPNRMQ 431
DB 337 NETLRVANIIGVFRAMTDVEIKTKIPGMVVFSSFRVHLDPNHFKDARTFNPWRM 366
QY 432 QONNGASSGSGSFTWGN--YMPFGGPRLCAGSELAKEMAVFIHVLKFNWELAE 490
DB 397 -----SNSVTIGSPSNVFTFPGGPRLCGYELARVALSVFLHRLVTGFSWPAEQ 446
QY 491 DQFFAPFVDFPGLIRVSR 511
DB 447 DKLVFPFTRTKRKYPIFKR 467

RESULT 5
AAG45021
ID AAG45021 standard; Protein; 491 AA.

AC AAG45021;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 56468.

KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay: genetic mapping; gene expression control; promoter;
KM termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134371.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136382.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

QY 491 DQFAFPFVDFPNCGLPRVSR 511
| : | | | |
Db 466 DKLVEFFPTTRQKRYPIFVKR 486

RESULT 6
AAG44570
ID AAG44570 standard; Protein; 492 AA.
XX
AC AAG44570;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

```

PR 13-AUG-1999; 990S-014684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149722.
PR 23-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 25-AUG-1999; 990S-0149930.
PR 26-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 30-AUG-1999; 990S-0151080.
PR 31-AUG-1999; 990S-0151303.
PR 01-SEP-1999; 990S-0151438.
PR 07-SEP-1999; 990S-0151930.
PR 10-SEP-1999; 990S-0152363.
PR 13-SEP-1999; 990S-0153070.
PR 15-SEP-1999; 990S-0153758.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154039.
PR 22-SEP-1999; 990S-0154779.
PR 23-SEP-1999; 990S-0155139.
PR 24-SEP-1999; 990S-0155486.
PR 28-SEP-1999; 990S-0155659.
PR 29-SEP-1999; 990S-0156458.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159293.
PR 14-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 18-OCT-1999; 990S-0159638.
PR 21-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 25-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 26-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 28-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 29-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

```

Query Match 36.5%; Score 978.5; DB 21; Length 492;
 Best Local Similarity 41.9%; Pred. No. 7e-78;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

QY 12 LLLPSSLLFLTLKRRNRKTRFNLPGKSGMPFLGRTIGYIKPYTATTLGDFMOQH 71
DB 27 LLLSSIAAGFL--LLLRTRRRMGGLPPGSLPLIGETFOIGAKYKTENPPEPTIDER 83
QY 72 VSKYGIKIRSNLFGPEPTIVSDAGLNRFIIIONERLEECSTPRISGIIIGKMSMLVYGD 131
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

DB 84 VARYGVFMTHLEGEPTIFSADEPTNRFVLONGKLFECSTPASICNLGKSHLLMKGS 143
QY 132 MHRDRSISLNFSLHALKRTLLKQDVRRHLLFYLDSSMOONSIFSADEAKKFTFNIMARH 191
DB 144 LHKMHSLLTMSFANSTITKDHMLDIDRLVRFNIDSSRVLL--MEKAKITFELTVQ 201
QY 192 IMSNDPGEETEDLQKRYVFMKGVASAPNLGCTAVHKAOSRATILNFIEKMERL 251
DB 202 IMSFDRG-EWSESLRKREYLLVIGFSLPLPLSTTYIKRAIQAR-----RKVAELT 252
QY 252 DIKEEDQEEEVTEDEAEKSKSDHYRKQRTDDLLGWVLKHSNLSTEQILLLISLFLA 311
DB 253 VVWKKRREEEEGAE-----RKKDMALAILAADGFEDEIVPIVALVLA 298
QY 312 GHETSSVALAIAFLFOACPAVEELREHELEIARAKKLGSELMWDYKKMDPTOCYI 371
DB 299 GYETITSTIMTANKFLFETPLALQKHEHEKIRAKKSD--SYSLSEMSDYKSMPTOCYV 356
QY 372 NETLRIGNVYRELHRRALDVRKGYDIPSGWKVLPVISAVHLDNSRYDQPLFNPWRQ 431
DB 357 NETLRVANIIGVFRRAMTDVEIKGYKIKMKVFSFRAVHLDNHRFDANTFNPWRQ 416
QY 432 QONNGASSSGSFSFTWGMN--YMPRGCGPRLCAGSELALEMAVPTIHLVLFKNWLEAD 490
DB 417 -----SNSVTTGPSNVFTPPGGGPRLCPGYELARVALSVFLHRLVTGFSVPAEQ 466
QY 491 DQFAFPFVDFPNCGLPIRYSR 511
DB 467 DLVFEPTTRQKRRPIFYK 487

RESULT 7
AA644572
ID AA64572 standard; Protein; 444 AA.
XX
AC AA64572;
XX
DE 18-OCT-2000 (first entry)
XX
KW Arabidopsis thaliana protein fragment SEQ ID NO: 55848.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125784.
PR 25-MAR-1999; 990S-0126284.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 16-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.

```

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147418.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148173.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152362.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

Query Match	35.3%;	Score 946;	DB 21;	Length 444;
Best Local Similarity	41.8%;	Pred. No. 4.6e-75;		
Matches 198; Conservative	77;	Mismatches 161;	Indels 38;	Gaps 7

RESULT 8
AAG45023
ID AAG45023 standard; Protein: 444 AA.

KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

PN	EP1033405-A2
----	--------------

25-FEB-2000; 2000EP-0301439.

PF 25-FEB-2000; 2000EP-0301439.

XX	25-FEB-1999;	9905-0121825
PR	05-MAR-1999;	9905-0123180
PR	09-MAR-1999;	9905-0123548
PR	23-MAR-1999;	9905-0125788
PR	25-MAR-1999;	9905-0126264
PR	29-MAR-1999;	9905-0126765
PR	01-APR-1999;	9905-0127462
PR	06-APR-1999;	9905-0128234
PR	08-APR-1999;	9905-0128744
PR	16-APR-1999;	9905-0129845
PR	19-APR-1999;	9905-0130049
PR	21-APR-1999;	9905-0130493
PR	23-APR-1999;	9905-0130510
PR	28-APR-1999;	9905-0130891
PR	30-APR-1999;	9905-0131449
PR	03-MAY-1999;	9905-0132048
PR	04-MAY-1999;	9905-0132407
PR	05-MAY-1999;	9905-0132485
PR	06-MAY-1999;	9905-0132486
PR	07-MAY-1999;	9905-0132633
PR	11-MAY-1999;	9905-0134266
PR	14-MAY-1999;	9905-0134218
PR	14-MAY-1999;	9905-0134219
PR	14-MAY-1999;	9905-0134221
PR	14-MAY-1999;	9905-0134370
PR	18-MAY-1999;	9905-0134368
PR	19-MAY-1999;	9905-0134941
PR	20-MAY-1999;	9905-0135141
PR	21-MAY-1999;	9905-0135333
PR	24-MAY-1999;	9905-0135629
PR	25-MAY-1999;	9905-0136021
PR	27-MAY-1999;	9905-0136382
PR	28-MAY-1999;	9905-0136782
PR	01-JUN-1999;	9905-0137222
PR	03-JUN-1999;	9905-0137528
PR	04-JUN-1999;	9905-0137502
PR	07-JUN-1999;	9905-0137724
PR	08-JUN-1999;	9905-0138094
PR	10-JUN-1999;	9905-0138540
PR	10-JUN-1999;	9905-0138847
PR	14-JUN-1999;	9905-0139119
PR	16-JUN-1999;	9905-0139452
PR	17-JUN-1999;	9905-0139463
PR	18-JUN-1999;	9905-0139464
PR	18-JUN-1999;	9905-0139465
PR	18-JUN-1999;	9905-0139456
PR	18-JUN-1999;	9905-0139457
PR	18-JUN-1999;	9905-0139458
PR	18-JUN-1999;	9905-0139459
PR	18-JUN-1999;	9905-0139460
PR	18-JUN-1999;	9905-0139461
PR	18-JUN-1999;	9905-0139462
PR	18-JUN-1999;	9905-0139463
PR	23-JUN-1999;	9905-0140535
PR	24-JUN-1999;	9905-0140695
PR	28-JUN-1999;	9905-0140823
PR	29-JUN-1999;	9905-0140921
PR	30-JUN-1999;	9905-0141287
PR	01-JUL-1999;	9905-0141842
PR	02-JUL-1999;	9905-0142154
PR	06-JUL-1999;	9905-0142905
PR	08-JUL-1999;	9905-0142390
PR	09-JUL-1999;	9905-0142820

PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 15-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144332.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155559.
 PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 946; DB 21; Length 444;
 Best Local Similarity 41.8%; Pred. No. 4.6e-75;
 Matches 198; Conservative 77; Mismatches 161; Indels 38; Gaps 7;

QY 39 LPKSGMPFLGEGITGVKPTATTLDPMOQHSKYKIRSNLFGEPVTSADAGNR 98
 DB 3 LPKSGMPFLGEGITGVKPTATTLDPMOQHSKYKIRSNLFGEPVTSADAGNR 62
 QY 99 FTLONEGRLEFECSTYRSIGILGKWSMLVLVDHMDRSTSLNLSHARLTILKDYE 158
 DB 63 FVLONEGRLEFECSTYRSIGILGKWSMLVLVDHMDRSTSLNLSHARLTILKDYE 122
 QY 159 RHTLFVDSWOQNSIFSAODEAKKFTFNMAKHHMSMDPGEETDOLKKEVTFMKGVYS 218
 DB 123 RHTLFVDSWOQNSIFSAODEAKKFTFNMAKHHMSMDPGEETDOLKKEVTFMKGVYS 179
 QY 219 APULPGAVYKKAALOSRATILKFTERKKEERKLDKEEDDEEVEKTEDEAEMSDHVR 278
 DB 180 APULPGAVYKKAALOSRATILKFTERKKEERKLDKEEDDEEVEKTEDEAEMSDHVR 219
 QY 279 KQRTDDDLGAVLNKSHNLSTEQIIDLILSLFAGHETSSVAIALAIFELQACPAVEELR 338
 DB 220 -RKMDLAALLAADDFSDDEIYDPLVALLVAGYETSTIMTLAVFLFETPLAQLK 277
 QY 339 EHELEIRAKKEGESELMNDYKMDOTQCVINETLRLGAVNFRFLHRAKLDKDYKCYD 398
 DB 278 EHELEIRAKKEGESELMNDYKMDOTQCVINETLRLGAVNFRFLHRAKLDKDYKCYD 335
 QY 399 IPGKWKVLPYSAVHLNRSRDQNLFPNPMWQOONNGASSGSGSFTMGNM-YMPRG 457
 DB 336 IPGKWKVLPYSAVHLNRSRDQNLFPNPMWQOONNGASSGSGSFTMGNM-YMPRG 385
 QY 458 GPRLCAGSELAKEMAVFIHHLVLYKFMWELAEDDQPAFPVDFPNCGLPIRVSR 511
 DB 386 GPRLCAGSELAKEMAVFIHHLVLYKFMWELAEDDQPAFPVDFPNCGLPIRVSR 439

PR	04-MAY-1999	9905-0132464
PR	05-MAY-1999	9905-0132465
PR	06-MAY-1999	9905-0132466
PR	07-MAY-1999	9905-0132467
PR	11-MAY-1999	9905-0132468
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134961
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137682
PR	03-JUN-1999	9905-0137558
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	14-JUN-1999	9905-0138847
PR	16-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	17-JUN-1999	9905-0139433
PR	18-JUN-1999	9905-0139432
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139889
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140655
PR	28-JUN-1999	9905-0140823
PR	30-JUN-1999	9905-0140991
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141482
PR	02-JUL-1999	9905-0142154
PR	06-JUL-1999	9905-0142305
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144532
PR	21-JUL-1999	9905-0144584
PR	21-JUL-1999	9905-0144614
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088


```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159658.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 27.1%; Score 726; DB 21; Length 461;
 Best Local Similarity 32.3%; Pred. No. 1.5e-55;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

```

QY 12 LLLPSLLSLLFLILLLKRRN--RKRFPMLPEKSGWPLGTTIGYIKPYATTLGDMQ 69
DB 7 LLIVSLCSALL-----RNMQWRRTKNGLPPTGMWPIFGTTFELKQGP-----NPMR 55
QY 70 QHVSXYKITYRSNLEGEPTIVSADAGLNFILIONEGRLFECSYPSIGIGLKMSMLIV 129
DB 56 NQRLRYGSPFKSHLGCPLILMSDSVNNYILINKSKGLVPGYPOSMIDLGTGCMVAHV 115
QY 130 GDMHRMRSISLNFSLHARLFTLLKDVRRHTLFVLDWQONSIFSADKAKFTF----- 185
DB 116 GSSHRLMRGSLLSLSTMTMRDHIPLPKVDHFMRSYLDQWNELEVIDDKTKHMAFSSL 175
QY 186 -----NIMAKHIMSMDPGEETEOLEKKEVTFMKGVASAPLNPETATKALOSRATILK 240
DB 176 TQIAGLTKRPY-----EEKTIAPFKLIVGTLSPIDLPFTNRCGIQARNIDR 225
QY 241 FIERMEERKLDIKEDDEEERKTEDEAEMSKSDHVRKORTDDILGMLKHSN----L 296
DB 226 LLRELMOERR-----DSGE-----TFDMGLYLLKKKGNRPL 258
QY 297 STEOILLILSLFNGHTSSVAIALAFPIQACPRAVEELREEHLEIARAKELGSEL 356
DB 259 TDEIRDOVVTILVSGYETVSTSMALRKYLDHPRKALQELRAHLEAFREKRQ--DEPL 316
QY 357 NMDDYKKMDFTQCVINETLRIGNVVRFLRRALKDQVRKYGKIDPSGKVLPISAVALDN 416
DB 317 GLEDVRSKMFTRAVIYETSRLATIVNGVLRKTTDLEINGLIRKGRATVYTRREINYDA 376
QY 417 SRVQPNLEFNPRMOQONNGASSSGSGSESTWGNMYPFGGGRPLCAGSELAKLEMAFVI 476
DB 377 NLEYDEPLLENPMRWKMSKLSLQ-----NSCFVFGGGRPLCGKELGIYEISFL 425
QY 477 HHLVLFKFWELAEEDOPFAFPFVDFPNGIPRVS 510
DB 426 HLFVTRRYRMEBIGDELMVFPFVAFKGFHRLIS 459

```

RESULT 12
 AAG46489 standard; Protein; 462 AA.
 AAG46489;
 18-OCT-2000 (first entry)
 Arabidopsis thaliana protein fragment SEQ ID NO: 58493.
 Protein identification: signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX

OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134258.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135114.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

	Query Match	27.18:	Score 726;	DB 21:	Length 462;
	Best Local Similarity	32.38:	Pred. No. 1.5e-55;		
	Matches 166;	Conservative	91;	Mismatches 181;	Indels 76; Gaps 10;
OY	12	LLLPPLSLILLFLILKRRN--RKTNRNPGRSGMFEGLGETTGLKPYATITLGDMDQ	69		
Dy	8	LIIIVISCSALL-----RWNQMYTKKGIDPPMGWVIFGEETEEELQGP----	NMR 56		
OY	70	QHWASYGKIYRSINLFGEPTIYSADAGINRLFIIONEGRLFECSPYSIGILGKMSLVAV	129		
Dy	57	NQLRYGSFPKSHLGCGPTLISMDSQEVNRYILKKNSKGLVGVCYQSMIDIIGTCMAAHH	116		
OY	130	GDMHRDQRSISLNFLSHARLTITLKDVERTHVLVDWSMOONSIFSADDEAKKETFT---	185		
Dy	117	GSSHRIMRGSLISTSIITMMRDHILPVDHFMRSYLDQWNELEVLDIDQTKRHNAFLSSL	176		
OY	186	-----NLMAKHINSMDPGSEETPDQLKEEYVFPMGVVSAPLNFGTAHYHKLQSATTLK	240		
Dy	177	TQTAGNLRKPVV-----EEFKFAFFKLVAGTLSPRIDLPGTNRCGIOARNNIDR	226		
OY	241	FIERKMERRLDIKEEDQEEREVEYTEDEAEKSKSDHYRKQRTDDDLIGWLXHSN----	L 296		
Dy	227	LLELMQGER-----DSGE-----TFMDMLIGLKKBSGNRYPL	259		
OY	297	STEOIDLILSLFAGHTSSVAIALAIFLQAOPKAVELREHELEIARAKKETGESSEL	356		
Dy	260	TDEIRIQOVTIITYSGYEIVSTSMMALKTYLHDHPKALQELRAAHIAHFRERKRQ--DEPL	317		
OY	357	NMDYKMDTQCVIINTLGLGVNVRPLHRRKALDKVARYKQDGIDPSGVKVFVISAVHLDN	416		

Db	318	GLEDVKSMTFAVIVETSRLATIYNGVLKRTKTDLEINCYLLPKGRITVYTRREIN	YDA 3777
QY	417	SKYDOPNLFNPPRWOOQNGASSSGSFSTWGNNTYMPGGPRLCAGSELATLEMAVFT	476
Db	378	NIYEDPLIFNPPRWKMSLSLSQ-----NSCFVGGGTRLCPGKELGIVEISSFL	426
QY	477	HLVLKFNWMLAEDDOPFAFPVPDPNGPLIRYS	510
Db	427	HYFVTRRWEETIGDELMVFPFVPAKGFHLRIS	460

RESULT 13			
AC	AAG46491		
XX	AAG46491	standard; Protein; 465 AA.	
XX	AAG46491;		
DT	18-OCT-2000	(first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 58495.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
XX	termination sequence.		
OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
XX	25-FEB-2000; 2000EP-0301439.		
XX			
PF	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0128845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131445.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	05-MAY-1999;	99US-0132484.	
PR	06-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	07-MAY-1999;	99US-0132487.	
PR	11-MAY-1999;	99US-0132863.	
PR	14-MAY-1999;	99US-0134236.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	

PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139859.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140981.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149928.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154019.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160771.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160960.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; Length 465;
 Best Local Similarity 32.3%; Pred. No. 1.5e-55;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
 12 LLLPILLSLLFLILKRRN--RKTRENLPPGKSGMPFGETIGLYKPYTATTLGDFMQ 69
 QY

[illegible]

PR	30-APR-1999	9905-0132048
PR	04-MAY-1999	9905-0132447
PR	05-MAY-1999	9905-0132445
PR	06-MAY-1999	9905-0132446
PR	07-MAY-1999	9905-0132463
PR	11-MAY-1999	9905-0132687
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134221
PR	16-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134944
PR	21-MAY-1999	9905-0135194
PR	25-MAY-1999	9905-0135639
PR	27-MAY-1999	9905-0136021
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139181
PR	16-JUN-1999	9905-0139422
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139869
PR	23-JUN-1999	9905-0140053
PR	23-JUN-1999	9905-0140054
PR	24-JUN-1999	9905-0140095
PR	28-JUN-1999	9905-0140082
PR	29-JUN-1999	9905-0140091
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141464
PR	01-JUL-1999	9905-0141254
PR	01-JUL-1999	9905-0142051
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0143377
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144353
PR	21-JUL-1999	9905-0144384
PR	21-JUL-1999	9905-0144414

PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 13-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 26.6%; Score 712; DB 21; Length 461;
 Best Local Similarity 32.5%; Pred. No. 2,5e-54;
 Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

QY 12 LLLPSSLILFLILKRRN--RKTRENLPPGSGMPFLGTTGYLKPYATYTGDMQ 69
 Db 1 MMILGLVIVICLTALRWNNOMRYSKGLPPTGTMPIFGTTEFLKGP-----DPMK 55
 QY 70 QHVSRYGKIVRSNLEGEPTIVSADAGLNRFILONGRLFECSYPSIGIGLGMMLVIV 129
 Db 56 NORLRYGSEFESHILGCPITIVSMDLELNRYILMNSKGLVAGYPOSMIDLITGTCINAVH 115
 QY 130 GDMHRDMRSISLNFSLHARLRTILKLDVERHTLFLVDSMOONSIFSADEAKKFTFNMA 189
 Db 116 GPSHRLMGSLLSISPTPMKDHLLPKTIDDFRNVLGMDLLETVDIEKTKHMAF-LSS 174
 QY 190 KHMSMDGEETEOLEKKEYTFPMKGVVSAPLNPFGTYHKALOSRAATILKFTIERKMEER 249
 Db 175 LLOIAETLLKKEVEEYRTEFFKLVVGLTSPIDIPGTNYRSGVQARNIDRLTLEIMOER 234
 QY 250 KLDIKKEQDEEEVYKTEDEAMSKSDHVRKQRTDDLLGVLKHSN---LSTEQILDLI 305
 Db 235 K-----ESGEFTT-----DMIGYLMKKEDNNRYLTDKRIROV 267
 QY 306 LSLPAGHETSSVAIALAIFFLQACPKAVEELREHLEIARAKKELGSELMNDYKMD 365
 Db 268 VTILYSGYEVSTTSMALKYLDHPKALEELREHLEIARERKR--DEPLTDDIKSMK 325
 QY 366 FTQCYINETLRLGNVVRRLHAKALKADVRRKGYDIPSGKVLPIVISAHLDSRYDQPIULF 425
 Db 326 FTRAVIFETSRILATVNGVLKRTYHDLNLNGYLIPKGRIRIYVYTRREINVDTSLYEDPMIF 385
 QY 426 NPMRMOONNGASSSGSFSFTWCNNYMPFGGPGPLGAGSELAKEMAVFIHHLVKRNW 485
 Db 386 NPMRMEKSLSKS-----YFLDFGGVRLCCKGKELGISVSLHLYFTYTKRW 434
 QY 486 ELAEDDOPFAFPVDFPNGLPRIYS 510
 Db 435 EENGEDKLMVFPVRSAPRGYHLKCS 459

RESULT 15
 AAG11835
 ID AAG11835 standard; Protein: 462 AA.
 XX AAG11835;
 AC AAG11835;
 XX 17-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 10707.
 XX Protein identification: signal transduction pathway; metabolic pathway;
 DE hydriatation assay; genetic mapping; gene expression control; promoter;
 KW

termination sequence.
KM XX Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0127485.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131431.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134368.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135553.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140355.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147439.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0156559.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 26.6%; Score 712; DB 21; Length 462;
 Best local Similarity 32.5%; Pred. No. 2.5e-54;
 Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

QY 12 LLLPLSLLLPLLLKRRN--RKTFRNLPKSGMPLGFTIGYLPYTATTLGDPMQ 69
 DB 2 MMLIGLVIIVICLCTALLRNNOMKYSKGLPPGTGMPRIFGETTEFLKGP----DFMK 56
 QY 70 QHVSXGKIYRSNLFGEPTIVSADAGLNRFLLONEGRFLPFCOSYRSIGILGKSMELIV 129
 DB 57 NQRLRIGSFKSHLGCPTIVSMDAELNRIILMNESGLVAGYPOSMDLIGTCNTIAVH 116
 QY 130 GDMHRDMRSISLNFSLHARLRTLLKDYERHTLFLDSWOONSIFSADAEAKKFTFNMA 189
 DB 117 GFSHRLMRGSLSLISPTMKDHLPRIDPFMRNYLCGMDLETVDIQEKTHAF--LSS 175
 QY 190 KHTMSMDPGEETEDQLKEVYPMKGVASAPLNPCTAYHKALOSRAITILKFTERKMEER 249
 DB 176 LLOIAETLKKRPEVEEYTFEFLVGTLSVPIDIPGTNYSRGVQARNNIDRLLELMQER 235
 QY 250 KLDIKEDEEVEYKTEDEAEMSKSDHVRKQRTDDDLGWLKHSN----LSTEQILDLI 305
 DB 236 K-----ESGETFT-----DMLGYLMKKEDNRNYLLTDKEIRDOY 268
 QY 306 LSLFAGHETSSVAIALAIFFLQACPAVEELREHLEIARAKKESEIMNDYKMD 365
 DB 269 VTILYSGYEIVTSMMALKYLLHDHPRALBELREHAIIRKRKP--DEPLTLDIDIKSMK 326

QY 366 FTQCVINETLRIGNVVRFLHRRKALADVRYKGTIDISGMKVLPVISAHLNDSRYDQPNLF 425
 DB 327 FTFAVIFETSRLATIVNGVLRKRTHTDLIELNGYLIPKGRIVYVYREINVDTSLEYEDPMIF 386
 QY 426 NPMRMQOONNGASSSGSFSFTWGNKNYMPFGGSPRLCAGSELAKEMLAVFIHHLVLEKFW 485
 DB 387 NPMRMKESLESK-----YFLIFGGVRLCPKEKELGISEVSSFLHYFVTKYRW 435
 QY 486 ELAEDDQPFAPFVDFPNGLPIRVS 510
 DB 436 EENGEDKLMPFPRVSAPKGYHLKCS 460

Search completed: June 10, 2003, 21:57:23
 Job time : 46 secs

• •

• •

```

?      ATTORNEY/AGENT INFORMATION:
?      NAME: Hunt, John C.
?      REGISTRATION NUMBER: 36,424
?      REFERENCE/DOCKET NUMBER: 50767/00004
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (416) 863-4344
?      TELEFAX: (416) 863-2653
?      INFORMATION FOR SEQ ID NO: 2:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 492 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?
US-08-724-466B-2

```

Query Match	18.8%;	Score 503;	DB 3;	Length 492;
Best Local Similarity	27.6%;	Pred. NO. 1.5e-41;		
Matches 144; Conservative	95;	Mismatches 216;	Indels 66;	Gaps 14

[illegible]

RESULT 4
 US-08-164d-2
 ; Sequence 2, Application US/08882164D
 ; Patent No. 6306624
 ; GENERAL INFORMATION:
 ; APPLICANT: Petkovich, P. Martin, White, Jay A.,
 ; TITLE OF INVENTION: Retinoid Metabolizing Protein
 ; NUMBER OF SEQUENCES: 43
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Blake, Cassels & Graydon
 ; STREET: Box 25, Commerce Court West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5L 1A9
 ; COMPILED READABLE FORM:

```

1 MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
2
3 COMPUTER: COMPAQ, IBM PC compatible
4
5 OPERATING SYSTEM: MS-DOS 5.1
6
7 SOFTWARE: WORD PERFECT
8
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/882,164D
11 FILING DATE: June 25, 1997
12
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: 08/667,546
15 FILING DATE: June 21, 1996
16
17 APPLICATION NUMBER: 08/724,466
18 FILING DATE: October 1, 1996
19
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Hunt, John C.
22
23 REGISTRATION NUMBER: 36,424
24 REFERENCE/DOCKET NUMBER: 50767/00010
25
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (416) 863-4344
28 TELEFAX: (416) 863-2653
29
30 INFORMATION FOR SEQ ID NO: 2:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 492 amino acids
33 TYPE: amino acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36
37 US-08-882-164D-2

```

Query Match	18.8%;	Score 503;	DB 4;	Length 492;
Best Local Similarity	27.6%;	Pred. No. 1.5e-41;		
Matches 144;	Conservative 95;	Mismatches 216;	Indels 66;	Gaps 14;

QY	HTLLPILPLPSLLFLF-----ILKRNKRKTFPNLPKSGSPGLFGLIGVLR	58
Db	4 YTLMTVECTLYLPIVLLFLAAKYLEMIMIRVDPNCSPILPBGMLPFIPTGLIOLI--	61
QY	59 YVATVLGDMOQVHSKYGIYRSLNLFGEPTIVSADAGLNREILONEGRLEFCSYPSRIGG	118
Db	62 --LQRKKLIRMKRQKUYGCGIYKTHLFGNPTRVAMGADVNRQILLEGHKLVSQNPASVRT	118
QY	119 ILCKSMVLVGDMDHRD-----MRSLSNPLSHARLRTILLKQDVRHTLPLVDSM--QONS	172
Db	119 ILGSDPLSNVGHQHKRKKALMRAPSRDALEH-----YIPVLIQOEKYSALQIOMLOKDS	172
QY	173 IFSADCAKFTFNLMAKHIMSDGEBET--EOLKEEYTFMKGVVSAPLNPFGTAHYK	230
Db	173 CVLVPEMKKKLMFRAMRILLGFEPEQIKTDEQELVEAFEMEIKNPLSLPIDVPSGLYR	232
QY	231 ALQSRATILKFLERKMEERKLDIKEODEEVEVLTDEAEKSDHNRKQRTDDLLGWY	290
Db	233 GLRARR----NFIHSHKIEENIRKKIDDQDNEHQKMLQOL-----IENSRSRE-----	279
QY	291 LKHSNLSTEQILLDLISLFAGHETSVAVALAIFLQACPAKAVEELREHLELTARAKE	350
Db	280 ----PFSIQAMEKATELLFGHEHTTASTATSLVMFLGMLNEVQYKREE----YQKKE	331
QY	351 LG-----ESGLNDDQKKKMDFTQCVYINELRLGNNVYRFLHKALKADVKYKGTIDPSGKVL	406
Db	332 MGAATPRGKLSMELLDOIKYTGCVAKELRLRINPVRGGFRVALKTFEFLNGYQIPIKGNVVI	391
QY	407 PVIASVHLDSNRKYDQNLFNFPWMQOONNAGSSGSGSFSTWGNVYMPFGGGRILGSGE	466
Db	392 YSICDPHDVADVFRKKEEOPEREH-----SKGLDGSNF--NYIPFGGSRMCGYKE	442
QY	467 LAKLEMAVFIHHLVLKFWMLAEDDO-----PFAFPRVDP	502
Db	443 FAKVLLKTFVLVELQHCWMLISNBPRTWKGTPTIYRVDNLP	483

RESULT 5
US-08-882-164D-32
; Sequence 32, Application. US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,16AD
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-16AD-32

Query Match 17.2%; Score 461; DB 4; Length 497;
Best Local Similarity 28.4%; Pred. No. 2.4e-37;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 10 LPLLLPSL-----LSLLFLLLK-----RRNRKTRFNLPPKSGMPLFGTIGYLK 57
DB 3 LPLALLASLCTFVLPPLLLFLAALKMDLYCVSSHRSCALPLPGTMGFPFGFTL---- 58
QY 58 PYATATLGDMMQOHVSXKGIYRSNLGEPTIVSADAGLNFTIIONEGRLFECSYPSIG 117
DB 59 -QWVLORRKFLQMKRRKTYGTYKTHLFGRPVVRWGANVRRILLGHRILVSVHWPASVR 117
QY 118 GILGKWSMLVYVGDHMDRMSISLNFSLHARLFTLLKDVERRHTLVLDSSWQ--NSIFS 175
DB 118 TITGAGCLSNLHDSHOKKVKIMQAFSREALQCYLV-IAEYSSCLEQWLSGEGGL 176
QY 176 AODBAKFTTNLAKKHIMSDFE---EETEDQKKRYVFMKGVASAPLNLPGTAYHKA 231
DB 177 VVEYEVKRLMFRIMRILGCEPAGGEGDEQQLVEAFEEETRLNLSPLIDVPESGLYRG 236
QY 232 LOSRATLTKFERKME--RKLDIKEED---QEEEVKTEDEAMSDHVKRKTDD 286
DB 237 VKRNLHARIEENIRAKIRRLQATEPDPGCKDALQLLIHSW-----RGERLDMQ- 288
QY 287 LGWVLKHSNLSTBOILDLLSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
DB 289 ---ALKOS---STE-----LFGGHETASATSLITVLGLYPHVLQVREETKSKGL 335
QY 347 AKELGESENLMDYKMDFTQCVINETTLRGVNVRELHRRKALKDVYKYGYDIPSGVKYL 406
DB 336 LKASNDNRKLDMETLEQLKYIGVYIKETRLNPNPVGFRVALKTFELNGLQIPKGNVI 395
QY 407 PVTASVHLDNSRDPDQPLNPPMRMOOONNASSSGSSSTWGNNTYPPGCGPRLCAGSE 466

DB 396 YSICDTHVDADIFTNKEEFNDPRTVPHPEDASRFS-----FIFGGGLRSCVCKE 446
QY 467 LAKLEMAVFIHLVLFKFMNELAE-----DDQFAPFPVDFP 502
DB 447 FAKILKIFTVELARHCWQMLNGPPTKTSPTVYPVNDLP 487

RESULT 6
US-08-724-466B-4
Sequence 4, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-4

Query Match 16.4%; Score 439; DB 3; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLLLK-----RRNRKTRFNLPPKSGMPLFGTIGYLK 57
DB 3 LPLALLASLCTFVLPPLLLFLAALKMDLYCVSSHRSCALPLPGTMGFPFGFTL---- 58
QY 58 PYATATLGDMMQOHVSXKGIYRSNLGEPTIVSADAGLNFTIIONEGRLFECSYPSIG 117
DB 59 -QWVLORRKFLQMKRRKTYGTYKTHLFGRPVVRWGANVRRILLGHRILVSVHWPASVR 117
QY 118 GILGKWSMLVYVGDHMDRMSISLNFSLHARLFTLLKDVERRHTLVLDSSWQ--NS 172
DB 118 TITGAGCLSNLHDSHOKKVKIMQAFSREALQCYLV-IAEYSSCLEQWLSGEGGL 176
QY 176 AODBAKFTTNLAKKHIMSDFE---EETEDQKKRYVFMKGVASAPLNLPGTAYHKA 231
DB 177 VVEYEVKRLMFRIMRILGCEPAGGEGDEQQLVEAFEEETRLNLSPLIDVPESGLYRG 236
QY 232 LOSRATLTKFERKME--RKLDIKEED---QEEEVKTEDEAMSDHVKRKTDD 286
DB 237 VKRNLHARIEENIRAKIRRLQATEPDPGCKDALQLLIHSW-----RGERLDMQ- 288
QY 287 LGWVLKHSNLSTBOILDLLSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
DB 289 ---ALKOS---STE-----LFGGHETASATSLITVLGLYPHVLQVREETKSKGL 335
QY 347 AKELGESENLMDYKMDFTQCVINETTLRGVNVRELHRRKALKDVYKYGYDIPSGVKYL 406
DB 336 LKASNDNRKLDMETLEQLKYIGVYIKETRLNPNPVGFRVALKTFELNGLQIPKGNVI 395
QY 407 PVTASVHLDNSRDPDQPLNPPMRMOOONNASSSGSSSTWGNNTYPPGCGPRLCAGSE 466

```

Db      286 DMO-----ALKOS--STE-----LLFGHETTASATSLITYLGLYHVLQVREBK 331
QY      343 EIAARAKKEGESELMNDYKMDPTQCVINETRLGNVVFELHRAKALDVKRYGYDIPSG 402
      332 SKGLLCKSNODNKLDMELIQLKYGICVIKETRLNPPVGGFRVALKTFELNGYQIPKG 391
QY      403 WKVLPYISAVHLNDSRYDQPNLFNPMWOOQNNGASSSGSFTWGNMMPGGGPRLC 462
      392 WNYIYSICDTHDVAEIFTNKEEFPDRFSAPHEDASRFS-----FIFPGGLRSC 442
QY      463 AGSELAKELEMAVFIHHLVLFKNWELAE-----DDQFAPFPVDFP 502
      443 VGKEFAKILIKIFTVELARHCWQLNGPPTKTSPTVYVDNLP 487

```

RESULT 7

```

US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-4

```

```

Query Match      16.4%; Score 439; DB 4; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;
QY      10 LPULLPSL-----ISLLFLILK-----RRNRKTRFNLPPKSGMPLGEGTIGYLK 57
      3 LPALLASLCTFVPLPLFLAIKIMLDYCVSGRDRSCALPLPCTMGFPFGEGL-----58
QY      58 PYATATLGDPMQOAHRSKYGIYRSNLGEPTIYASADALNFIIONBRLTECSPPRSIG 117
      59 -QWVLRKRFIQMRKRTGFIYKTHLFRPTVRWAGADNVRKILLGDRLVSVHRPASVR 117
QY      118 GILGKWSMLVVLGDMHRMRSISINFLSHARL---TLLKDVERTHLFVLDMSQO--NS 172

```

```

Db      118 TILSGGSLNLDHSSHQRRKVIKRAFSREALCYVVIIEVGSS-----LEQWLSGGER 173
QY      173 IFSAOEAKKFTFNIAKHIKMSDP-----GEBETDOLKEVYTPMKGVSAALNPGTA 227
      174 GLVYEVKRIKLMRIAMRILLGCEPOLAGGDSG--QOLVAFEFEMTRNLFSLPIDVFGS 232
QY      228 YHIALDSRATILKFERKMEERLIDKEED-----OEEBKTEDEAKMSDHYRKORT 282
      233 LYRGKARNLIHRIQINIKICGLRASAAGGCCDAQQLLHESME-----RGERL 285
QY      283 DDLLDMVLKHSNLSEQLIDLILLSLFAGHESVAIALATFFLOACRKADELREBH 342
      286 DMO-----ALKOS--STE-----LLFGHETTASATSLITYLGLYHVLQVREBK 331
QY      343 EIAARAKKEGESELMNDYKMDPTQCVINETRLGNVVFELHRAKALDVKRYGYDIPSG 402
      332 SKGLLCKSNODNKLDMELIQLKYGICVIKETRLNPPVGGFRVALKTFELNGYQIPKG 391
QY      403 WKVLPYISAVHLNDSRYDQPNLFNPMWOOQNNGASSSGSFTWGNMMPGGGPRLC 462
      392 WNYIYSICDTHDVAEIFTNKEEFPDRFSAPHEDASRFS-----FIFPGGLRSC 442
QY      463 AGSELAKELEMAVFIHHLVLFKNWELAE-----DDQFAPFPVDFP 502
      443 VGKEFAKILIKIFTVELARHCWQLNGPPTKTSPTVYVDNLP 487

```

RESULT 8

```

US-09-615-192A-405
; Sequence 405, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,769
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 405
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-615-192A-405

```

```

Query Match      11.3%; Score 302.5; DB 4; Length 529;
Best Local Similarity 21.9%; Pred. No. 1.7e-21;
Matches 117; Conservative 104; Mismatches 204; Indels 109; Gaps 18;
QY      11 PULLPSLISLLFLILKRRNRKTRFNLPPKSGMPLGEGTIGYKPYATATLGDPMQO 70
      22 PFTLLSVPLLLFLGLVARLRKRPPF--PGPRLPLVIGNML-----MMGEILHR 70
QY      71 HVS-----KYGKIYRSNLGEPTIYASADAGLRFTIIONBRLTE-----CSYRSI 116
      71 GLASLAKYGGIFILRMQFLIMVAVSSDVAROVLOHDCGIFSRRPATAISYLYPDAD 130
QY      117 GGIIGKWSMLVVLGDMHRMRSIS-LNFLSHARLRTILKDVERTHLFVLDMSQONSIFS 175
      131 -----MAFAHYGPFWRMRKLCVMKLFSSKR-----AESWE-----S 162
QY      176 AODEAK-----KFTFNIAKHIKMSDPGHEETEOLAKKEVYTPK-----214
      163 VRDEVDMVRIVAGSEGTAVNGELVLFELTRDIYRAAFGTSSTEG--ODEFISILQBSK 221

```


QY 169 QONSTFSAODEAKKFTFNIMAKHIMSDPEEETDOLKKEVTFMGVSA----- 219
DB 174 KNGGVWVWLKDAFAVFCILVAMCFGLMEDEETVERIDQ---VMSVLITLPRIDDL 229
QY 220 PLNLPR--GATYHALOSRATILKFERKMEERKLIDKEEQEEVEKTEDEAKMSDVH 277
DB 230 PILSPFSKORKKALVREQVEFLVITEQRRRAIONPG-----SDHT 273
QY 278 RKORTDDL-----GWLKHSNLSSTEOIIDLILSLFAGHETSSVALAIFLQACP 331
DB 274 ATTFSLDILFLDLKAGV---KKSAPDAELVSLCSEFLNGTDTTAAEWGIAQLTANP 330
QY 332 KAVEELREHELEIARAKKEGESELMWDDYKKMDFTQCVINETILRLGNVVRFLHRAKLD 391
DB 331 NVOTKLYEE-----IKRTVGEKKVDEKDEKMPYLHAYVKELRHHPHFVLTAAVTE 384
QY 392 -VRKYGDIPSGKVLVPSAVHLNDSRIDQPNLFNPMWQOONNGASSGSGSFSTWGN 450
DB 385 PTTGGYDIPIDANVEVYTPALAEDEKRNMLNPEKPERFERISGGEADITG-----VTGV 439
QY 451 NYMPFGGPRLCAGSELAKLEMAVFIHLVLKFNW 485
DB 440 KMPFSGVGRICGFLMAIVHILMAKRVQEFEM 474

RESULT 11
US-09-144-367-2
; Sequence 2, Application US/09144367
; Patent No. 6432639
; GENERAL INFORMATION:
; APPLICANT: Lichter, Jay
; APPLICANT: Guido, Marco
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/09/144,367
; CURRENT FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/058,612
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: H. sapiens
; US-09-144-367-2

Query Match 10.2%; Score 273; DB 4; Length 503;
Best Local Similarity 22.4%; Pred. No. 1.4e-18;
Matches 120; Conservative 101; Mismatches 222; Indels 92; Gaps 21;
QY 9 LILPILSLILLLILILKRRNRKTRFNLPR---PGKSGMPLEGTIGYLKPYATTL 64
DB 3 LIPDLAMETWLLAVSLVILYLGTHSHGLFKLGLIPPLPFLGNIILSYHKGFCM--- 59
QY 65 GDFMOQVSKYKGIYSNLEGEPTIVSADAGLNRFLIONGRLEFESYPSIGILG--K 122
DB 60 --FDMCHKKYKGVWFYDGOQPVLAITPDMIKTVLKE--CYSFTNRPFGPYGFMK 115
QY 123 WSLVLVGDHMDRMS--ISLNLFSH-----AKRLITLKDVERH-----TLFVL 165
DB 116 SAISIADEEMKRLRLSLPTFSGLKLEKVPILAYGYDLVRLNLRREAFGTGPRVTL-- 172
QY 166 DSMQONSIFSA--QDAKKFTFNIMAKHIMS--MDPGEETEOLK--EYV--TEKKGVYSA 220
DB 173 -----KDYFGAISMVYITSTSGVINDSLNNPDPREVENTKLLRDFDLDFLSTIVP 227
QY 221 LNPSTAYKALQSRAITILKFERKMEERKLIDKEDEEVEKTEDEAKMSKSDHYRK 280
DB 228 FLIPILLEVILNICVFPREVTNFL--RKSVKRMRKESRLBDQKHRY--DFQL--MIDGNSK 282
QY 281 RKDDDLGWLKHSNSTEOIIDLILSLFAGHETSSVALAIFLQACPRAVEELRE 340

DB 283 ETE-----SHKALSDLEIVAOSIIFIPAGYETSSVLSFIWYELATHPDVOOKLOE 334
QY 341 HLEIARAKKEGESELMWDDYKKMDFTQCVINETILRLGNVVRFLHRAKLDVRKGYDIP 400
DB 335 -IDAVLPNK-----APPLYDVLOWEYIDVYNETLRLPLAMLERCKKDVEINGMFI 389
QY 401 SGKVLVPSAVHLNDSRYQPNLFNPMWQOONNGASSGSGSFSTWGNP-----YMPF 455
DB 390 KGVWVMIPSYALHNDPRYWEPEKFLPERSSKKK-----DNIDYIITPF 435
QY 456 GGGPRCAGSELAKLEMAVFIHLVLKFNWELADDPPEAPVDFPNGLPIRYS 510
DB 436 GSGFRNCIGMRFALMNMKLLIRVLQNFSEPKCKETO-----IPLKLS 478

RESULT 12
US-09-126-420A-18
; Sequence 18, Application US/09126420A
; Patent No. 6376753
; GENERAL INFORMATION:
; APPLICANT: BATARD, YANNICK
; APPLICANT: ROBINEAU, TIBURCE
; APPLICANT: DURST, FRANCIS
; APPLICANT: WERCK-REICHAUT, DANIELE
; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP6B1 FROM HELIANTHUS
; TITLE OF INVENTION: TUBEROSUS AND ITS DEGRADATION OF ENVIRONMENTAL
; TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
; TITLE OF INVENTION: POLYUNTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
; FILE REFERENCE: 03715.0032
; CURRENT APPLICATION NUMBER: US/09/126,420A
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 60/054,351
; PRIOR FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Solanum melongena
; US-09-126-420A-18

Query Match 10.1%; Score 271; DB 4; Length 504;
Best Local Similarity 21.7%; Pred. No. 2.2e-18;
Matches 111; Conservative 95; Mismatches 231; Indels 74; Gaps 17;
QY 12 LILPILSLILLLILILKRRNRKTRFNLPRPGKSGMPLEGTIGY-LKPYATTLGDFPMQ 70
DB 13 IILIPA-----FILFSSQKNTTKSSYRPPGPGPLPIFGNMFEIGTEPYKMAV---LRQ 63
QY 71 HVSYKGIKYSNLEGEPTIVSADAGLNRFLIONGRLE-----ECSYRSITGI 119
DB 64 ---KGFVLMKESSTYTMVVOYTRQASEEELKNDISFANRVIPDVNOAHSYOGSLA-- 118
QY 120 LGRWMSLVYVGDHMDRMSI--SLNPLSHARLITLKDVERHTEFLVDSMQONSIFSAOD 178
DB 119 -----IAPYGPMPFRQRICTIEMFHHKISE--TEPVARRKCDNMNLKWEKANAER 170
QY 179 -----EAKF-----TFNIMAKHIMSMDPEEETDOLKKEVTFMK-----GVSAFPLNPG 225
DB 171 GSGIEYTRFVFLASFNMGLILSKDLADLSESEASFEFTAMKRINMSGIANVSDITFP 230
QY 226 TAYHKAQSRAITILKFERKMEERKLIDKEDEEVEKTEDEAKMSKSHVROKPTDD 285
DB 231 LKRFDLQSLKKKARDKGAVELIMSFLKE--KEEKKKGTGKGLDVLVLEFQGTGKD 288
QY 286 LIGWVLKHSNLSSTEOIIDLILSLFAGHETSSVALAIFLQACPRAVEELREHELEIA 345
DB 289 -----EPAKLSHEIRIIFVLEMLAGETETSSSEVALLELLRHPRAMAKVTEISQAI 342
QY 346 RAKKEGESELMWDDYKKMDFTQCVINETILRLGNVVRFL--HRAKLDVRKGYDIPSGMK 404
DB 343 EPNRKEDS-----DIENLPYMOAVILKESRLHPRPLPFLIPRETIOPTKMGYDVPRQTO 397

QY 405 VLPVISAHLNDSRYDQNLFPNRMQOONNGASSSGSFSTWGNNT--MPFGGPRLC 462
 Db 398 VLVNAMAIGRDEPCMDPMSEKPERFL-----GSKIDVAGOHGLIPFAGRRMC 447
 QY 463 AGSELAKLEMAVFIHHLVLFKFNWELAEDDP 493
 Db 448 VGLPLGHRMHFALGSLREFEWELPGVSP 478

RESULT 13

US-09-499-302A-2
 ; Sequence 2, Application US/09499302A
 ; Patent No. 6369212
 ; GENERAL INFORMATION:
 ; APPLICANT: BOUNG-JUN, OH
 ; APPLICANT: MOON, KYUNG KO
 ; APPLICANT: YOUNG, SOON KIM
 ; TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
 ; TITLE OF INVENTION: INCOMPATIBLE INTERACTION
 ; FILE REFERENCE: 10324/P6443050
 ; CURRENT APPLICATION NUMBER: US/09/499, 302A
 ; CURRENT FILING DATE: 2000-02-07
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 502
 ; TYPE: PRT
 ; ORGANISM: Capsicum annuum
 ; US-09-499-302A-2

Query Match 10.0%; Score 268.5; DB 4; Length 502;
 Best Local Similarity 21.5%; Pred. No. 3.9e-18;
 Matches 114; Conservative 100; Mismatches 208; Indels 109; Gaps 18;

QY 17 SLISLLFL--ILKKRRNKRTRBNLPKSGMPELEETIGYLPYATITLGD-----MQ 69
 Db 7 NLVAFLLFSLIILLKKMKTKQLDPPGKLPFISL-----FLAVAGPLPHGLK 60
 QY 70 QHVSQKIRSNLFGPTIVSADAGLNFTLQNEGRLEFCSYPRSIGIL-----120
 Db 61 QLAKLYPLMLHRLGEIPYIISPRMAKVKLTHDLAFNRFKLVADIVHYOSTDIAF 120
 QY 121 ---GKSMVLVGDHMDMSISLNFSLHARLTILKDVRRHFLVYDS---WQONSIF 174
 Db 121 SPYGEWT-----RQIRICILELISAKM-VKFFSSIRQDELISMVVSIRTPNRPV 170
 QY 175 SADEAKKFTFNLMAKHMSMDPEEETDQKKRYTFMKGVVSAP-----220
 Db 171 NLDDKITWFTSSVYCRSALGKICNDQ-----KLITFMREITISLTGGFSTADFPPTWK 224
 QY 221 LNLPGTAUHKALOSRAATILFKIERKMEERKIDIKEEDQEEBEVTEDEAE---MSKSDH 276
 Db 225 LHDVGGSKTRLKAKHRIDELIEHVNVENHKONRADGQKGGEGGEDLDVLLAVRESGE 284
 QY 277 VRKORTDDLLGWLLKSNLSTEDQILDLSLTFAGHETSSVALALAIFFLQACPKAVEE 336
 Db 285 VQISTTDN-----IKSILYDMFSAGSETSTIIMALAEKKRPSYLAK 329
 QY 337 LREHELEIARAKKELESELMWDDYKKMDFOCYINTEFLRGVNVRL-HRKAKLDVRYK 395
 Db 330 AQAEVROYLKEKK--GRQQLDLDELK--YIKLVIKETLRNHPRIPLLVPECKKDTYID 384
 QY 396 GYDIPSGMYLPLVISAHLNDSRYDQNLFPNRMQOONNGASSSGSFSTWGNNTMPF 455
 Db 385 GYNIPFETRIYVNMALIGRDEPCMDPMSEKPERF--ENSSVDLGS-----HHQIFTF 436
 QY 456 GGGPRLCAGSELAKLEMAVFIHHLVLFKFNWELAEDDPFAFPVDFPNGLP 506
 Db 437 GAGRRICPG-----MKFG--LANVGGPLAQLLYHNFRIKL 469

RESULT 14

US-09-126-420A-17
 ; Sequence 17, Application US/09126420A
 ; Patent No. 6376753
 ; GENERAL INFORMATION:
 ; APPLICANT: BATARD, YANNICK
 ; APPLICANT: ROBINEAU, TIBURCE
 ; APPLICANT: DURST, FRANCIS
 ; APPLICANT: WERCK-REICHAUT, DANIELE
 ; APPLICANT: DIDIERJEAN, LUC
 ; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
 ; TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
 ; TITLE OF INVENTION: PARTICULAR FOR THE DEREGULATION OF ENVIRONMENTAL
 ; TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
 ; FILE REFERENCE: 03715.0032
 ; CURRENT APPLICATION NUMBER: US/09/126,420A
 ; CURRENT FILING DATE: 1998-07-30
 ; PRIOR APPLICATION NUMBER: 60/054,351
 ; PRIOR FILING DATE: 1997-07-31
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 17
 ; LENGTH: 467
 ; TYPE: PRT
 ; ORGANISM: Solanum melongena
 ; US-09-126-420A-17

Query Match 10.0%; Score 268; DB 4; Length 467;
 Best Local Similarity 23.2%; Pred. No. 3.9e-18;
 Matches 113; Conservative 80; Mismatches 182; Indels 112; Gaps 18;

QY 46 WPLIGETIGYLPYATITLGDPMQOH-VSKYKIRSNLFGPTIVSADAGLNFTLQNE 104
 Db 28 WLKIGSSMNTVLTQANASSELFRNHVS-----FSDRPYDVNLAHNY-----72
 QY 105 GRLEFCYSRISIGILGKSM-LVLVGDHMDMSI-SLNFSLHARL-RTILKDYERHT 161
 Db 73 -----KSMALAPGYMFRSRICVEMVHNRINETTIRQESVDK 115
 QY 162 LVLDSMOONSIFSAD--EAKKF---TFNLMAKHMSMDPEEETDQKKRYTFMKG 215
 Db 116 MLRLDEEKASSGGGEGIEVTRYVFLASFNNVGMIFSKDLVTPDSKOGSEFFNAMIG 175
 QY 216 VVSAPLNLPGTA-----YHKAQD-----SKATILFKIERKMEERKIDKEED 257
 Db 176 IMEW-AGVPNLSIDIFPCIKMFDVQGLRKKMERDMGKRETKKFTLEERKEK-----227
 QY 258 QEEBEVKTEDEAKMSKSDHVRKORTDDLLGVL-----KHSNLSLSTEQILDLSL 309
 Db 228 -----KGE-----KNRSIKRLDVLIDFEGSGKDEPKLSEDEITVITLEKF 269
 QY 310 FAGHETSSVALALAIFFLQACPKAVEELREHELEIARAKKELESELMWDDYKKMDFQC 369
 Db 270 LAGETWSSVEMALTELLRHRQAMAKVLEITQYIGPNKKFEEDCID--SLPYQA 324
 QY 370 VINETLRGNVNVRL-HRKAKLDVRYKGYDIPSGKMYLPLVISAHLNDSRYDQNLFPN 428
 Db 325 VLKEDQLRHPRLPLIPKKAIDQKKMEYDIPKGYVLVNMALIGRDEPYWNPPEEFKE 384
 QY 429 RMOOONNGASSSGSFSTWGNNT--MPFGGPRLCAGSELAKLEMAVFIHHLVLFKFNWE 486
 Db 385 RFLF-----SKVYVKGQNTYELIFFGAGRRRCVGLPLGRHMHNTTFGSLHREFWE 434
 QY 487 LAEDDP 493
 Db 435 LPHNVSP 441

RESULT 15

US-09-126-420A-26
 ; Sequence 26, Application US/09126420A
 ; Patent No. 6376753
 ; GENERAL INFORMATION:
 ; APPLICANT: BATARD, YANNICK

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:57:53 ; Search time 32 Seconds

(without alignments)
1655.071 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681
Sequence: 1 MFETHTLPLLLPSLLS.....FAFPVDFPNCPIRVSRL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep: *
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep: *
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep: *
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep: *
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep: *
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep: *
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep: *
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep: *
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep: *
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep: *
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep: *
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep: *
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep: *
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	816.5	30.5	486	9	US-10-142-231-92 Sequence 92, Appl
2	597.5	22.3	485	9	US-10-142-231-88 Sequence 88, Appl
3	596.5	22.2	497	9	US-10-142-231-64 Sequence 64, Appl
4	592	22.1	496	9	US-10-142-231-90 Sequence 90, Appl
5	588	21.9	509	9	US-10-142-231-67 Sequence 67, Appl
6	585	21.8	507	9	US-10-142-231-65 Sequence 65, Appl
7	583	21.7	484	9	US-10-142-231-57 Sequence 57, Appl
8	571	21.3	501	9	US-10-142-231-58 Sequence 58, Appl
9	569.5	21.2	493	9	US-10-142-231-89 Sequence 59, Appl
10	568.5	21.2	500	9	US-10-142-231-56 Sequence 56, Appl
11	564	21.0	484	9	US-10-142-231-61 Sequence 61, Appl
12	564	21.0	500	9	US-10-142-231-62 Sequence 62, Appl
13	558.5	20.8	483	9	US-10-142-231-87 Sequence 87, Appl
14	558.5	20.8	512	9	US-10-142-231-87 Sequence 87, Appl
15	553.5	20.6	498	9	US-10-142-231-63 Sequence 63, Appl
16	534.5	19.9	503	9	US-10-142-231-60 Sequence 60, Appl
17	458	17.1	492	9	US-10-210-965-4 Sequence 4, Appl
18	302.5	11.3	491	9	US-10-067-534-3 Sequence 3, Appl
19	292	10.9	516	9	US-10-067-534-3 Sequence 3, Appl

20	290.5	10.8	514	10	US-09-947-027-4 Sequence 4, Appl
21	290.5	10.8	514	12	US-10-091-009-4 Sequence 4, Appl
22	289	10.6	511	10	US-09-796-256a-4 Sequence 4, Appl
23	285.5	10.6	502	9	US-10-054-988-69 Sequence 69, Appl
24	285.5	10.6	502	10	US-09-739-254-69 Sequence 69, Appl
25	285.5	10.6	502	10	US-09-804-615-69 Sequence 69, Appl
26	275	10.3	496	9	US-10-067-668-10 Sequence 10, Appl
27	275	10.3	496	9	US-10-175-696-10 Sequence 10, Appl
28	275	10.3	496	9	US-09-945-301-4 Sequence 4, Appl
29	273	10.2	503	10	US-10-146-575-2 Sequence 2, Appl
30	272	10.1	470	9	US-10-067-668-12 Sequence 12, Appl
31	272	10.1	470	9	US-10-175-696-12 Sequence 12, Appl
32	264	9.8	512	9	US-10-142-231-66 Sequence 66, Appl
33	262.5	9.8	512	9	US-10-201-213-5 Sequence 2, Appl
34	258.5	9.6	574	9	US-10-236-433-2 Sequence 2, Appl
35	258	9.6	514	9	US-10-142-231-68 Sequence 2, Appl
36	256.5	9.6	520	10	US-09-992-901-2 Sequence 2, Appl
37	251	9.4	515	10	US-09-796-138-19 Sequence 19, Appl
38	251	9.4	515	10	US-09-909-903-19 Sequence 19, Appl
39	251	9.4	544	9	US-10-067-668-8 Sequence 8, Appl
40	251	9.4	544	9	US-10-175-696-8 Sequence 8, Appl
41	249	9.3	503	10	US-09-796-138-18 Sequence 18, Appl
42	249	9.3	503	10	US-09-909-903-18 Sequence 18, Appl
43	249	9.3	503	10	US-09-957-997-3 Sequence 3, Appl
44	249	9.3	509	10	US-09-817-184-4 Sequence 4, Appl
45	247.5	9.2	493	9	US-10-103-520-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1	US-10-142-231-92	Sequence 92, Application US/10142231
Publication No.	US20030077796A1	
GENERAL INFORMATION:		
APPLICANT:	Croteau, Rodney et al.	
TITLE OF INVENTION:	CYTOCHROME P450 OXYGENASES AND THEIR USES	
FILE REFERENCE:	62773	
CURRENT APPLICATION NUMBER:	US/10/142,231	
PRIOR FILING DATE:	2002-05-08	
PRIOR APPLICATION NUMBER:	60/165,250	
NUMBER OF SEQ ID NOS:	95	
SOFTWARE:	PatentIn Ver. 2.1	
SEQ ID NO 92		
LENGTH:	486	
TYPE:	PRT	
ORGANISM:	Taxus cuspidata	
US-10-142-231-92		
Query Match	30.5%; Score 816.5; DB 9; Length 486;	
Best Local Similarity	35.1%; Pred. No. 2.2e+58;	
Matches	176; Conservative 97; Mismatches 184; Indels 45; Gaps 8;	
QY	19	LSLLFLILKRRNRK-----TFNLPFGKSGMPTGCTIGYKPYATATLGPMDQHV 72
Db	20	LFICVILLILRRNSDROGNSANPKILPGSAGLPFGITRTLRDASKPGRKFFDEHE 79
QY	73	SKYKTIYSNLFGEF-TIVSDAGLNFRTILIONBGRLEFCSTYPSRISGIGLQMSMLVLVD 131
Db	80	LRYGPIFCSTLGFRTAVSVDPFNKYVLONBGRLEFSNALAFRNILIGYGLSAVOGE 139
QY	132	MHRDRSISLNFSLHARLRTILKDVSRHTLFVLDVSDMOONSIFSAODEAKKFTNNLAKH 191
Db	140	LQRLHTFAYVLLKHLFLSSDFMEDIDITQAGRKKEEGDIPIDKCNQVYINLMAKR 199
QY	192	IMSNDPEEETEQKKEYVTFMGVASAPLNDGTAYHKALOSRAATLTKFERKMERKL 251
Db	200	LLDLPP-SEMGHGYKAFDFGVAVLSFPINPGTGYARIGIRAKGLIKKHICIERR- 257
QY	252	DIKEEDQEEEVATEDEAEMSKSDHVRKAKTDDDLGLGVYKHSNLTGQILDLITSLPLA 311

```

Db 258 -----BHEPV-----LRNDLTKLVREGTSDLEIADTIFEVPA 292
Qy 312 GHETSSVAIALAIFLQACPAVEELREEHLEIARAKKEGSELNMDYKMDPTQCVI 371
Db 293 GVEISAMAMTAVKVLAKENPRALEHDLAKAKG-GNEKLTWMDYQSMKRVHCYI 351
Qy 372 NETRLGAVNVPFLHKKALKDVRKYDIPSGKVLVPSAVHLDSRDOPLFNPWRQ 431
Db 352 NETRLGATVYVLFREAKQDITKVDYLPKGMTVSVFSAHVVDKRYHEADKFLPWRQ 411
Qy 432 QONNGASSSGSFSTWGNNTMPFGCGPRLCAGSELAKEMAVEITHLVKFNWELAEED 491
Db 412 NE-----GQETLEECYMPFEGRGRLCPGLALFELALFHLNVTYKRMQLEID 462
Qy 492 QPFAFPVDFPNGPLPIRV-SRI 512
Db 463 RAYFPPLPSTENGPEIRLYSRV 484

RESULT 2
US-10-142-231-88
; Sequence 88, Application US/10142231
; Publication No. US20030077796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 485
; TYPE: PRF
; ORGANISM: Taxus cuspidata
US-10-142-231-88

Query Match 22.3%; Score 597.5; DB 9; Length 485;
Best Local Similarity 27.9%; Pred. No. 1.6e-40;
Matches 147; Conservative 114; Mismatches 191; Indels 75; Gaps 13;

Qy 9 LLLPILLPSL-----LSLLFLPILLLKRRNRKTRFNLPPKSGMPLGCTIGLKPYATTATL 64
Db 4 LKOLEVSPSILFVTLAVAGIILFFRSKRHSVXKLPNGLPPLVGETLQEVRSIGSTP 63
Qy 65 GDFMQOHVSKYKTYRSNLFGEPTVSADAGLNRFILQNGRLEFECSPRSIGILGKWS 124
Db 64 QQFTEERMSKRGDVKFTSIIGHPTVVLGCPAGNRLVLSNENKLVOMSPSSMKLLIGEDC 123
Qy 125 MLVIVGDMHRDMSISLNFSLSHARLTILK--DVERHTLFLVDSMQ-----Q 170
Db 124 LGKRTGQHRIYRAALTRFLGRQALQNHFAKMSGIOH--INEKMGKDEAVPLVYK 180
Qy 171 NSISADDEKAKFTFNMMAKIMSMDPEEETEOLKKEYVTFMKGVSAPLNPGLTAYHK 230
Db 181 DLVFSV--ASRLFGTTEHLDQ-----EQDHLNLEVLVSGFSVPLINIGFSYKH 228
Qy 231 ALQSRATILKFIERKMEERKLDIKEDEEEVKTDEDAEMSKSDHVRKQRTDDLLGMV 290
Db 229 ALQKRALDADMTILIKRRRELKAGTASENQ-----DLSVL 266
Qy 291 L-----KHSNLEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELREEHLEIA 345
Db 267 LTFDERGNSLADKEILDNFSMLHGSYDSTNSPLMLIKVLASHPSYEVKVADEQGITL 326
Qy 346 RAKKEGESELNMDYKMDPTQCVINETRLGAVNVPFLHKKALKDVRKYDIPSGKVL 405
Db 327 STKME-GE-EIAMDILKEMKYSWQVQETLRMPPIGTGTFKRAITDIHNGYTIPIKWKL 384
Qy 406 LPVLSAVHLDSRDOPLFNPWRMOQONNGASSSGSFSFTWGNNTMPFGCGPRLCAGS 465

```

```

Db 385 LMTTSTQTKKEYFKADQPKPSRFEDEGKHNP-----YTYLPPFGGMVCPGW 434
Qy 466 ELAKEMAVEIHLVLKFNWELAEED--DQPAFPVDFP-NGLPIRV 509
Db 435 EFARKMETLFLHHPVAFSGKALIDPNEKLSGRPLRPLVNGPLRITL 481

RESULT 3
US-10-142-231-64
; Sequence 64, Application US/10142231
; Publication No. US20030077796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 497
; TYPE: PRF
; ORGANISM: Taxus cuspidata
US-10-142-231-64

Query Match 22.2%; Score 596.5; DB 9; Length 497;
Best Local Similarity 29.4%; Pred. No. 2e-40;
Matches 152; Conservative 98; Mismatches 200; Indels 67; Gaps 12;

Qy 16 PSLSLLL-----FLILKRRNRKTRFNLPPKSGMPLGCTIGLKPYATTATLQDFMQ 69
Db 21 PALSLTLAPLAIILLLLFRRYNNHRSVXKLPNGLPPLVGETLQILFRLRSPTPKFRD 80
Qy 70 QHVSRYKTYRSNLFGEPTVSADAGLNRFILQNGRLEFECSPRSIGILGKMSLVLY 129
Db 81 DLKRFEGPYMTSLGHPTVVLGCPAGNRLVLSNEKLVEMGPSFEMKLIQEDSIVARR 140
Qy 130 GDMHRDMSISLNFSLSHARLTIL--LKDVERHTLFLVDSMQONSIFSAODEAK----- 181
Db 141 GEDHILKRALARFLGAGALQNYLGRMSEIGHN-----NEKMGKDEAVKVLVY 191
Qy 182 -KFTENLMAKHIMSMDPEEETEOLKKEYVTFMKGVSAPLNPGLTAYHKALQSRATYIK 240
Db 192 RLQISISTLFEDVNDGHQ-Q-KQLHLLLETLVGSLSVLPDPRGRYRKGLQARLKDE 250
Qy 241 FIERMEERKLDIKEDEEEVKTDEDAEMSKSDHVRKQRTDDLLGMV-----KHSN 295
Db 251 ILSLTKRRRDLKSGIASD-----DQDLSVLTLFPRDEKGS 288
Qy 296 LSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELREEHLEIARAKKEGESE 355
Db 289 LITDQIILNFSAMFASIDTYAPRALFLKLLYSNPEYENKFFQDQLEITGKKKE-GE-E 346
Qy 356 LNMDDYKMDPTQCVINETRLGAVNVPFLHKKALKDVRKYDIPSGKVLVPSAVHL 415
Db 347 ISWKDLKSMKYTWQVQESLSLRMPVPGIFKRAITDIHNGYTIPIKGMVILCSPYTTTH 406
Qy 416 NSRYQPLVFNWRMOQONNGASSSGSFSFTWGNNTMPFGCGPRLCAGSELAKEMAVE 475
Db 407 EEPYEPPEPRPSREDEGRHVP-----YTYVPFGGLRTCPGMEFSKLEILLF 456
Qy 476 IHLVLKFNWELAEED--DQPAFPVDFP-NGLPIRV 509
Db 457 VHHFVKNSSYIIPVDPRNEKVLSDPLPLPANGFSIKL 493

RESULT 4
US-10-142-231-90
; Sequence 90, Application US/10142231
; Publication No. US20030077796A1
; GENERAL INFORMATION:

```

Wed Jun 11 10:32:35 2003

us-09-502-426b-2.rapb

Page 3

```

: APPLICANT: Croteau, Rodney et al
: TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
: FILE REFERENCE: 62/773
: CURRENT APPLICATION NUMBER: US/10/142,231
: CURRENT FILING DATE: 2002-05-08
: PRIOR APPLICATION NUMBER: 60/165,250
: PRIOR FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 90
: LENGTH: 496
: TYPE: PRT
: ORGANISM: Taxus cuspidata
: US-10-142-231-90

```

Query Match 22.1%; Score 592; DB 9; Length 496;
Best Local Similarity 30.6%; Pred. No. 4.6e+40;
Matches 154; Conservative 93; Mismatches 215; Indels 42; Gaps 11;

QY	12	ILLLPSSLILLLFILLKRRKRTRFNLPEKSGMPLGEGTIGYIKPYATTLGD-FMQQ	70
Db	29	ITAPLPLSTL--RLKTAARRMP--LPPSGMMPWGESLIGSGMNNOSNDPVDYT	83
QY	71	HVSKYQKIYRNLGEGEPTIVSADAGLNRFILQNEGRLEFECYSPRSIGILGKWSMLVYG	130
Db	84	RKAKHGFIETHIGSPYVMLGPDANRFILINNKLEFLNSWPRSLNALKIKHALITSQG	143
QY	131	DMHRDMSISLNLPSHAKRLTILLKDYERHTLEVLD--WOONSIFSQDEAKKFTFNMA	189
Db	144	AEKRRMRILHSYGLPRNPETSVGR-FEGVLHLHLDSDMHGQIILQYRQYKDALCLAA	202
QY	190	KHMSMDGEEETEOULKEEYVTFMKGVSAPLNLPGAYHKLSDRAITIKFIERKMEER	249
Db	203	DFPMGLKPG-KELETFRRHSDPSAGILSHPLDPLWVFGKAKARAAVYQITSQRLH	261
QY	250	KLDIKEDDQEEEVKTEDEAEMSKSDHYRKQRTDDILLGVLKHNSLSTQDILLDILSL	309
Db	262	KFSMHSGEGGNFLDMVLGISOEKGDL-----LSEEEIADNLGMILL	304
QY	310	FAGHETSSVAIALAIFELQACPAVEELREEHLEIARAKKELGESLNMDYKMKMDTQC	369
Db	305	TGGDDTASALATILKHLSPILLQRLRKE-CEKLRDNKEAG-GPLTWSEITSVGYLHN	362
QY	370	VINETLRIGVAVFLHRKALKDYRYGYDIPSGMKVLVISAVHLNDSRYDOPLNFPMWR	429
Db	363	VISEGIRMAVPIINGEKKAKVDVYGGYITPKMKMKVAYSRYOTNNKEEYPPSPRFPDR	422
QY	430	WQOQNNGASSGSGSEFTGNMNPFGGSGRLGASLELAEMAVFTHHIYIKFENWELAE	489
Db	423	FNRHREPS-----FIPPGQGNRMKPGNEFAKLEMBELIYHLVLRKYDWELME	469
QY	490	DQGFAPFPVDPF-NGLPIRVSKRI	512
Db	470	ADERINMYFIPPHVHSLPLLLKIV	493

RESULT 5
US-10-142-231-67
Sequence 67, Application US/10142231
Publication No. US20030077796A1
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
FILE REFERENCE: 62773
CURRENT APPLICATION NUMBER: US/10/142,231
CURRENT FILING DATE: 2002-05-08
PRIOR APPLICATION NUMBER: 60/165,250
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 67
LENGTH: 509
TYPE: PRT

; ORGANISM: *Taxus cuspidata*
US-10-142-231-67

Query Match	21.9%;	Score 588;	DB 9;	Length 509;
Best Local Similarity	29.5%;	Pred. No. 1e-39;		
Matches 144;	Conservative 93;	Mismatches 189;	Indels 62;	Gaps 8;

QY	1	LLPSSLILLLLLLKRNKRTNRNPPGSGMPLFETIGYLRPTATTLGDPMQOHV	72
Db	24	IYLSAVAGIVLPLLLFKRSKRSSVGLRPPGKIGYFFIGESLFLFKALRSNTVDEFLDERV	83
QY	73	SKYGYLRNMLGPEPTIVSADAGILNRFILIONEGRLFECSYPSRISGIGLGMKSMVLVGM	132
Db	84	KNEGAVFKTSLIGHPLTVLCPGAPNRILLANEKULYQMSPKSMKLMGKSTIARKEG	143
QY	133	HHDMRSISLNFISHARKFI---LKDVRTELLVLDSDMOONSIFFSQO-----DEAK	181
Db	144	HHIINSALOGFSPALOKYIGOMKSTIENH---INEMKNDQSVSYVALVGLVDYDISA	200
QY	182	KETFLIMAKHNSMDPGEETEOLKKEVYTPMKGVSNAPLPLFTAAHKLQSRATITKF	241
Db	201	CLEFNNINEKH-----ERERLELEETIAYGVLAVPDLPBFAHNRALQARKSYMAI	251
QY	242	TERKREERLKDKEEDQEEBEVKYEDENEMKSHVARKQRTDDLLGWL-----KHSNL	296
Db	252	LSGLTERKMDLSS-----GLATSNDDLVSFLTFKFDKDGNGC	289
QY	297	STEOILDLILLFLFAGHETSSVAIALAIFLQACPRAVEELREHELETIARAKKEGSEL	356
Db	290	SDEELIDNFSGLHGSYDPTVSAMACVYKLSSNPECYERKVVYQDGLISNKLE--GDEL	347
QY	357	MNDQVKKMDQOCVINEFLRLGNVYRFLRKALADVKKGYDIPSGKVLVYISAVHLDN	416
Db	348	TKADKYSKMYTMQVQVETLRILYPSIFSGSRALTDIHYNYXIIPKQKMLMPEYTHNPKR	407
QY	417	SKYDQPNLFENWRMOONNGNAGSGSGSFTWGNANNYPARGGPRLCAGSELLEMAVFI	476
Db	408	MYFSEPEKFLPSRFQDEGKLVAP-----YTLPLPFGGGRSCPGMFESKMEILLSY	457
QY	477	HHLYLVKFN	484
Db	458	HHFVATFS	465

```

RESULT 6
US-10-142-231-65
: Sequence 65, Application US/10142231
: Publication NO. US20030077796A1
: GENERAL INFORMATION:
: APPLICANT: Croteau, Rodney et al.
: TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
: FILE REFERENCE: 62773
: CURRENT APPLICATION NUMBER: US/10/142,231
: CURRENT FILING DATE: 2002-05-08
: PRIOR APPLICATION NUMBER: 60/165,250
: PRIOR FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 65
:
: LENGTH: 507
: TYPE: PR1
: ORGANISM: Taxus cuspidata
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(507)
: OTHER INFORMATION: "Xaa" equals any peptide
:
: US-10-142-231-65

```

```

Query Match      21.8%; Score 585; DB 9; Length 507;
Best Local Similarity 30.6%; Prod. No. 1.8e-39;
Matches 158; Conservative 94; Mismatches 204; Indels 60; Gaps 14;

10 LPLLPLPSLLSLLFLLLRNRNRKTRFNLPPKSGMPLGTIGLYLKPYATVTLGDMQ 69

```

```

Db      14 LPEFLTP-ILGALVLLHFRHRNRSS-VKLPFGKLGPEVIGETIOFLALRSOPPOKFEF 71
QY      70 QHVSXKYIYRNLGEPPTIVSADAGLNRFILQNEGRLEFECYSRISIGILGKSMVLV 129
Db      72 DRYQKFGGVRTSLIGNPVLVCMGPAGNRVLVSNEDKLYOLEARNLSLMLKMGOSLAKR 131
QY      130 GDMHRDKRSISLNFSLHARLTILKDKVERHTLVLDWQONSIFSAODEAKKTFENLMA 189
Db      132 QEDHRTLRALARLGLFOALXNYMTKISSRTEHHMNEKMK-----GKDEVYR--TLPLIR 183
QY      190 KHMMSMDPG-----DEETDQLKEKYVTPEKGVVSAPLNPGTAHKAQSRATILKFI 242
Db      184 ELIFSNASSLFPDINDHQERLHLLEAVVGSMSIPLDPGTRLRALQARSLDEIL 243
QY      243 ERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDDLGWL-----KHSNLS 297
Db      244 SSLKSRKRL-----VSGIASD-----DQDLISVLTFRKDERGNFLT 281
QY      298 TEQILDILILSLFAGHETSSVALAIFLQACPKAVEELREHELEIARAKKEGESELN 357
Db      282 DKEILDNFSLLHASYDTVSPVLTLLKLSNDECEKYYVOEOLGIY-ANKRIGE-EIS 339
QY      358 WDOYKKKDETCQVNETLRGLGNVRFLLHRKALKDVRKGYDIPSGWKVLPVISAHLDNS 417
Db      340 WKDLKAKVYTWQVQETLRMPPLGSGFRKAMVDIDGYTIPKGMMLMTTGTHLREE 399
QY      418 RYDOPNLFNPMWMOQNNGASSSGSFSTGNNYMPFGPRLCAGSELAKLEMAVFIH 477
Db      400 YFNBPFLKFRPSRFE-----DGRVTPY--TFIFPGGARGCPGHEFSKTEILFLPIH 448
QY      478 HVLKENMELAE--DQFAFPVDFP-NGLPIRV 510
Db      449 HFVTFESSYLPVDSNEKISADPEFPLPANGFSIKLS 484

```

RESULT 7

```

US-10-142-231-57
; Sequence 57, Application US/10142231
; Publication No. US2003007796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 57
; LENGTH: 484
; TYPE: PRF
; ORGANISM: Taxus cuspidata
US-10-142-231-57

```

```

Query Match      21.7%; Score 583; DB 9; Length 484;
Best Local Similarity 29.3%; Pred. No. 2,4e-39;
Matches 154; Conservative 94; Mismatches 209; Indels 68; Gaps 13;

```

```

QY      5 EHHTLPLRLLLSLTLFLIL--KRRNRKTRFNLPRKSGMPFLGETIGYLRKPTA 61
Db      4 EMDTFVQLSSPVLSLTFLILFLIFCSKQYRSLPLPGNMGFLPGETIA---LAS 59
QY      62 TLLGDFMOQVSKYKLYRNLGEPPTIVSADAGLNRFILQNEGRLEFECYSRISIGILG 121
Db      60 QTDKPFQDMKKKFKKTSILGHPTIVLGGSSGNFLLSNEKVLAMPFPNPPSSSKILG 119
QY      122 KMSMLVLDGMHRDMRSISLNFSLHARLTILK--DVERHTLVLDWQONSIFSAOD 178
Db      120 QDSVLGIGBEHRIVPRALARCGLQALQNVVSKMSSEIORH--INQKMKGKEVYKMLP 176
QY      179 EAKKFTNLMKHMMSMDPGEETDQLKEKYVTPEKGVVSAPLNPGTAHKAQSRATI 238

```

```

Db      177 LIRLSVFIASLFEGL-TDEQOQERLHLLLETVTGLLCIPLDPGTFRRALHARSKL 235
QY      239 LKFIERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDDLGWL-----KH 293
Db      236 DEIMSVIERRRNDR-----LGAASSODDLISVLTFRKDERG 273
QY      294 SNLSTEQILDILILSLFAGHETSSVALAIFLQACPKAVEELREHELEIARAKKEGE 353
Db      274 NPFADKEILDNFSFILHALYDTTISPLTLVFKLVSNEPECYENIAOOLEILRNKKD-GE 332
QY      354 SELMNDVKKKDFQCVNETLRGLGNVRFLLHRKALKDVRKGYDIPSGWKVLPVISAHV 413
Db      333 -DISWADLKMKYTWQAVOETLRMCPPYAGNFRKALTDIHGDTYIPKGMWILCSPTYTH 391
QY      414 LDNSRYDOPNLFNPMWMOQNNGASSSGSFSTGNNYMPFGPRLCAGSELAKLEMA 473
Db      392 SKEEFDPDEKFRPSRFEQGRDVA-----YFIFPGGGLRICPGHEFAKMEIL 441
QY      474 VFHHLVLKF-----NMELAEDDQFAFPVDFP-NGLPIRV 509
Db      442 VFMHFVKAFFSFIIVDPNEKISTD-----PLPSIPVNGFSINL 480

```

RESULT 8

```

US-10-142-231-58
; Sequence 58, Application US/10142231
; Publication No. US2003007796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 58
; LENGTH: 501
; TYPE: PRF
; ORGANISM: Taxus cuspidata
US-10-142-231-58

```

```

Query Match      21.3%; Score 571; DB 9; Length 501;
Best Local Similarity 29.2%; Pred. No. 2,4e-38;
Matches 150; Conservative 104; Mismatches 200; Indels 60; Gaps 14;

```

```

QY      12 LLLPLSLLSLFLILKRRNRKTRFNLPRKSGMPFLGETIGYLRKPTATLGDPMOQH 71
Db      28 LLSLAFTAVALIFVLLFRKSRPSTNPPGNGFPFGETIGYLRALRSESPPHMFDER 87
QY      72 VSKYGIYRNLGEPPTIVSADAGLNRFILQNEGRLEFECYSRISIGILGKSMVLVGD 131
Db      88 LKKEGRVYKTSILGHPTAVFGCPAGNRRTIYSEHLLVSSGNSVYKLVGOSIYTKIGE 147
QY      132 MHRDMRSISLNFSLHARLTILK--DVERHTLVLDWQONSIFSAODEAKKFT 184
Db      148 EHRIFLGLVNEFLGPHALOSTYPMKMSKIOENIKH-----WKGXDEVNMLPSIRQV 200
QY      185 FNLMAKHMMSMDPGEETDQLKEKYVTPEKGVVSAPLNPGTAHKAQSRATILKFI 244
Db      201 FSISSSLEFDIN-DEDOQOELKTLETTLVGTLSVLDIPSGNFRKALRARSKLDEILSR 259
QY      245 KMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDDLGWL-----KHSNLS 299
Db      260 LIESRRKDMRGIAS-----ISKNLISVLLAFKDERGNPLDVT 297
QY      300 QILDILILSLFAGHETSSVALAIF-FLQACPKAVEELREHELEIARAKKEGESELN 358
Db      298 EILDNFSMLHASYDT-TVSPVLCIFKLISNPECEKYYVOEOLGIILNKKD-GE-EKCM 354
QY      359 DDYKKMDFTCQVNETLRGLGNVRFLLHRKALKDVRKGYDIPSGWKVLPVISAHLNSR 418

```

Db 355 NDKAKKTYWQAAQETWRLPPPAFGSFRKYIADIHHDGYIIPGKAMATNYSRKEY 414
 QY 419 YDOPNLEPMWMOOONNGASSGSESTWNNYMPGGGPRCAGSELAKLEMAVRIHH 478
 Db 415 FEDPDMFKPSRF-----GDGKYVA-PYTFILPFGAGIRICPGMEFAKLEMLFIHH 463
 QY 479 LUKENWEALAE--DOPPAFPVDFP-NGLPPIRV 509
 Db 464 FVKNFSGYLPLDTEKERISGDPFPLPKNGFPIKL 497

RESULT 9

US-10-142-231-89
 ; Sequence 89, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIORITY APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 89
 ; LENGTH: 493
 ; TYPE: PRP
 ; ORGANISM: Taxus cuspidata
 US-10-142-231-89

Query Match 21.2%; Score 569.5; DB 9; Length 493;
 Best Local Similarity 28.9%; Pred. No. 3,1e-38;
 Matches 146; Conservative 101; Mismatches 207; Indels 51; Gaps 11;

QY 18 LLSLLFLILL--KRRNRKTRFNLPPGSGWPLGFTIGYLAQPYATTLGDMOQHVSKY 75
 Db 23 ILSTLTALLLEFFRRNRKSSHKLPPGNLGFPIGETIOFLRSLSRQTEFFEDERVKFR 82
 QY 76 GKIRSNLFGPEPTVADAGLNRFILONEGRLEPCSYPRSIGILGKMSMLVVGDMHRD 135
 Db 83 GPVFKTSLGAPVIVICGAGSRVLVSNEDKLVMESPSSLAKLGMENSTILYREEHRT 142
 QY 136 MNSISLNLISHALRTILLL--DVERHTLVLDSSWOONSIFSQAODEAKFTENLMAKHI 192
 Db 143 LRSALSRFLGPOLQTYIAKMSLEIRH--INEKKKKEVYVTLDLIRGLVSIASSLPE 199
 QY 193 MSMDPEEETEOLAKKEYVTFMGVVSAPLNLPGTAVHKALQSRATILKFTERKMEERKLD 252
 Db 200 FDIIN-DEPOERLHHLESLVAGSMAVRLDFPOTRRKAVEARSKLDEALHSLIKSRSD 258
 QY 253 IKEEDDEBEVTEDEAEMSKSDHVAKORTDDLLGMVLKHSN-----LSTEQILLILLS 307
 Db 259 LLS-----GKASSNODLLSVLLSFKDERGNPLRDEEILLDNFSL 296
 QY 308 LIFAGHETSSVAIALAIFLQACPKAVEELREHEHLEIARAKKELGSEELNMDYKKMDFT 367
 Db 297 ILHASVDTLISPVVLLKLSSNPECYDVVOGFOELIANKKR-GE-ELSMKRLKMKMTY 354
 QY 368 QCVINETLRLGNVVRFLHKKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDOPNLEFP 427
 Db 355 WGVVDETLLMFPPLPFGSFRKAMVDINYGDTIPKGIIVLMTYSTHVKEEYFNEPGKFRP 414
 QY 428 WRMOOONNGASSGSESTWGNMMPGGGPRLCAGSELAKLEMAVFIHHLVLRKNWEI 487
 Db 415 SREHDDGRVAP-----YTFLPFGGGLTCGMEFSKTEILLFIHHFVKTSYLL 464
 QY 488 AED--DOPPAFPVDFP-NGLPPIRV 509
 Db 465 PVDPMKISADPFPPPLPANGFSIKL 489

RESULT 10
 US-10-142-231-59
 ; Sequence 59, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 59
 ; LENGTH: 500
 ; TYPE: PRP
 ; ORGANISM: Taxus cuspidata
 US-10-142-231-59

Query Match 21.2%; Score 568.5; DB 9; Length 500;
 Best Local Similarity 28.9%; Pred. No. 3,9e-38;
 Matches 150; Conservative 103; Mismatches 195; Indels 71; Gaps 14;

QY 16 PSLSTLL-----FLLLRNRNRKTRFNLPPGSGWPLGFTIGYLAQPYATTLGDMQ 69
 Db 24 PAILSTALTAIAGIYLVITSKRSSLAKLPGLGPIGETIEFVKALRSOTLRQFVE 83
 QY 70 QHVSYKGIKIRSNLFGPEPTVADAGLNRFILONEGRLEPCSYPRSIGILGKMSMLV 129
 Db 84 EREGKFGVFKTSILGKPFVILCGPAENRVLVSEERKLVHSQAQIRLILGINSVAVKR 143
 QY 130 GDMHRDNRSSLSNPLSHARLRTILKD---VERHTLVLDSSWO---QNSIFS-----AOD 178
 Db 144 GDHRYVLRVALAGFLGAGLQYIGKMSALIRNH--INRKKGKDVENVLSLVRIYMD 200
 QY 179 EAKKFTENLMAKHIMSMDPEEETEOLAKKEYVTFMGVVSAPLNLPGTAVHKALQSRATI 238
 Db 201 NSAILFENIYDK-----ERKQOLHEILKTIASHFGIPLINPEFLRKMLGSLKR 251
 QY 239 LKFTERKMEERKLDIKEEDDEBEVTEDEAEMSKSDHVAKORTDDLLGMVLKHSN--- 295
 Db 252 KKLISLALKERKDELRS-----KLASSNODLLSVLLSFRBERG 289
 QY 296 --LSTEQILLILLSIFAGHETSSVAIALAIFLQACPKAVEELREHEHLEIARAKKELGE 353
 Db 290 KPLSDEAVLDNCRAMLDASTDITTSOKTLLKLSNPECFEKEYVOEOLIASNKE-GE 348
 QY 354 SELNMDYKKMDFTOCVINETLRLGNVVRFLHKKALKDVRKGYDIPSGMKVLPVISAHV 413
 Db 349 -ELTMDIKAMKKTWQVLAQSLNMLSPVFGTLKTKTMDINHDTYIPKGVVWYWTYSTH 407
 QY 414 LNSRITDOPNLEPMWMOOONNGASSGSESTWGNMMPGGGPRLCAGSELAKLEMA 473
 Db 408 QKDLYFKOPDKFMPMSREED-----GHLDAY--TFVPEGGGGRATCGMEYAKVEIL 457
 QY 474 VFIHHLVLRKNWEALAE--DOPPAFPVDFP-NGLPPIRV 509
 Db 458 LFLHFFVKAPSGYTPDPHERICGYVPLVPVKGFPKL 496

RESULT 11

US-10-142-231-56
 ; Sequence 56, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12

QY 122 KMSALVGDHMRKMSISLNFLSHARLTLLK---DVERHTFLVDSMOONSIFSAD 178
 Db 119 QSLGKTGQEHRIYRTALGRFLGPOLQNHVAKKSSDIQNH---INOKMGNDEYKVL 175
 QY 179 EAKKFTFNLMAKHIMSMDGCEETBOLKEEYTFMKGVYSAFLNPGYAVHKALOSRAFI 238
 Db 176 LIRNLVFSATSLFEGIN-DEHOERLHLLETTYMGAVCIPLAFPGSGFRKALQARSEL 234
 QY 239 LKFIKKKEERLIDKEEDOEDEEVKTEDEAMSKSDHVRKQRTDDLLGVNLKHSNST 298
 Db 235 DGLISLMKIRSDLR-----SGAASSQDILSVLITFKDRG-----NPLTD 277
 QY 299 EQLILILSLFAGHETSSVALAIFLQACPAVEELREHLEIARAKKEGSELNWD 358
 Db 278 KEILDNFVYLHGLDITLPLTLFKLMSNTECEYVVOLEIL-SHREKE-ETGW 335
 QY 359 DDYKMDFTOCVINEETLRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDR 418
 Db 336 KDLKSMKYTWQALQETELMPPYGNFRKALTDIHYDITLPGKMRVLCSPFTTHSNEY 395
 QY 419 YDQNLNFWRMQOONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFIH 478
 Db 396 FNEDEFRPSRFEQGGKVP-----YTFIPGGGRLICPGMEFAKTEMLFIHY 445
 QY 479 LVLENNELAD--DQPFAPFVDP-NGLPIRV 509
 Db 446 FVKTFSSTVPVDPNEKISADPLASFPVNGFSVKL 479

RESULT 14

US-10-142-231-87
 ; Sequence 87, Application US/10142231
 ; Publication No. US20030077796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 87
 ; LENGTH: 512
 ; TYPE: PRT
 ; ORGANISM: *Taxus cuspidata*
 US-10-142-231-87

Query Match 20.8%; Score 558.5; DB 9; Length 512;
 Best Local Similarity 27.9%; Pred. No. 2,6e-37;
 Matches 143; Conservative 108; Mismatches 210; Indels 51; Gaps 12;

QY 10 LPLILPLSLILLLILKRRNRKTRFNLPGKSGMPLEGITIGLYKRYTATTLGDEM 69
 Db 27 LSVTIVTAIAVITLTLVLRSPQSCVNLPPGKLGIPRIGITLQLOAFRSNRPOQFD 86
 QY 70 OHVSKYKTYRSLNREPTIVSADAGLNFIIONBGRLEFECYPSRISGILGKWSMLV 129
 Db 87 EOKKFGSVYKTSLLGDRTVVLCGPSGNRLILSNENKILEVASMPSISIKLIGEDSIAGN 146
 QY 130 GMDHMDRSISLNFSLHARLTLLK---DVERHTFLVDSMOONSIFSADQAKKFTFN 186
 Db 147 GEKHRLRAVNRVYIGPALQNYMAKMSSEIHH---MEKKKGAEQYKVLPLVAVENFS 203
 QY 187 LMAKHIMS-DEGEETBOLKEEYTFMKGVYSAFLNPGYAVHKALOSRAFIKFIERK 245
 Db 204 IATSLFPGVNDGER-ERLHDILTALAGVFSIPDPGTYRKALBARLIDKLVSSL 261
 QY 246 MEERLIDKEEDOEDEEVKTEDEAMSKSDHVRKQRTDDLLG-VYL---KHSMLSTEQ 300
 Db 262 IERRSDLRSGVAGSNE-----DLSVWLITPKDEEGNPLTDKE 299

QY 301 ILDLILSLFAGHETSSVALAIFLQACPAVEELREHLEIARAKKEGSELNWD 360
 Db 300 ILDNFSTLHASDYTTLSALTTLKLMSSSTECYHKNVVOELRLVSKKE-GE-ETLKD 357
 QY 361 YKMDFTOCVINEETLRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDRYD 420
 Db 358 LKDKYTWQVVOETLRMPPLGSGFRKALTDIHYDITLPGKMRVLTWYTSJHOREYFN 417
 QY 421 QPNLNFWRMQOONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFIH 480
 Db 418 EPEKFMPSRFEERHAP-----YTFIPFGGVCVTCPEWERSKQIILFLHYEV 467
 QY 481 LKFWN--ELAEDDQPFAPFVDP-NGLPIRV 509
 Db 468 KTESGYIPLDDEKVLGNPVPPLRANGFAIKL 499

RESULT 15

US-10-142-231-91
 ; Sequence 91, Application US/10142231
 ; Publication No. US20030077796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 91
 ; LENGTH: 498
 ; TYPE: PRT
 ; ORGANISM: *Taxus cuspidata*
 US-10-142-231-91

Query Match 20.6%; Score 553.5; DB 9; Length 498;
 Best Local Similarity 29.6%; Pred. No. 6.4e-37;
 Matches 140; Conservative 101; Mismatches 191; Indels 41; Gaps 10;

QY 17 SLILSLFLILKRRNRKTRFNLPGKSGMPLEGITIGLYKRYTATTLGDEM 76
 Db 30 SALAGILLLILFRKRSKSLTPGKLGIPRIGISFIFLRALRNSLSQPFDERYKKG 89
 QY 77 KIYRNLGEPPIVSADAGLNFIIONBGRLEFECYPSRISGILGKWSMLV 136
 Db 90 LVEKTSLLGHPTVYVLCGPAGNRLILSNENKILEVQSWPQPMKMGENSEVATRRGDHYM 149
 QY 137 RSLNLFSLHARLTLLK---DVERHTFLVDSMOONSIFSADQAKKFTFN 193
 Db 150 RSLAGTFGPALQSYIGKMNTEIQSH---INERKKGDEVNVVLPVLELNFNLSALIF 206
 QY 194 SMDPGEETBOLKEEYTFMKGVYSAFLNPGYAVHKALOSRAFIKFIERKMEERKDI 253
 Db 207 NI-YDKQODRLHLKLETLIVGSFALPIDLGFGRHRLQGRALINKLMSLIRKRED 265
 QY 254 KEEDEEBEVKTE--DEAMSKSDHVRKQRTDDLLGVNLKHSNSTEQIIDLILSLFA 311
 Db 266 SLDRQOPRISLCSLSEMTKG-----LPH---PMDEILDNFSILHA 306
 QY 312 GHETSSVALAIFLQACPAVEELREHLEIARAKKEGSELNWDYKMDFTOCYI 371
 Db 307 SYDTTSPMALIFLLSSNRCYKQVVOLEIL-SHREKE-ETWMDLAKMYTWQVA 364
 QY 372 NETLRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDRYDQPNLFWRMQ 431
 Db 365 QETLRMPPLVPGFGRKAITIDYDGTNSKRG-KLMTTYSSTHPRDLYFNEPEKFMPSRFD 423
 QY 432 QONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFIH 484
 Db 424 QEGKHVAP-----YTFIPFGGQORSCVGEWERSKMETLFLVHHFVTFPS 466

Wed Jun 11 10:32:35 2003

Search completed: June 10, 2003, 22:03:35
Job time : 33 secs

us-09-502-426b-2.rapb

Page 8

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:57:28 ; Search time 162 Seconds

(without alignments)
2041.655 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681
Sequence: 1 MFEHTHTLPLLLPSLLS.....FAFPVDFPMLPIRVSRL 513

Scoring table:

BLOSUM62

GAPOP 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues 4569144

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

```
1: /cgn2_6/ptodata/1/paa/PCTUS.COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097.COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US098.COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US099.COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US100.COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US101.COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US102.COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60.COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	1	PCT-US00-03820-2
2	2681	100.0	513	19	US-09-502-426-2
3	2677	99.9	513	21	US-09-791-537-67246
4	1500	55.9	392	18	US-09-464-535-36
5	1016	37.9	484	21	US-09-708-427-51512
6	1016	37.9	485	21	US-09-708-427-51511

7	1016	37.9	542	21	US-09-708-427-51510	Sequence 51510, A
8	978.5	36.5	472	17	US-09-306-844-2	Sequence 2, Appl1
9	978.5	36.5	472	19	US-09-513-996A-55847	Sequence 55847, A
10	978.5	36.5	472	19	US-09-513-996A-55649	Sequence 56459, A
11	978.5	36.5	472	21	US-09-791-537-145750	Sequence 145750, A
12	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
13	978.5	36.5	472	23	US-09-935-625-5735	Sequence 5735, Ap
14	978.5	36.5	472	23	US-09-935-625-7739	Sequence 7739, Ap
15	978.5	36.5	472	23	US-09-935-625-9349	Sequence 9349, Ap
16	978.5	36.5	491	19	US-09-513-996A-55468	Sequence 55468, A
17	978.5	36.5	492	19	US-09-513-996A-55846	Sequence 55846, A
18	978.5	36.5	492	23	US-09-935-625-5226	Sequence 5226, Ap
19	978.5	36.5	492	23	US-09-935-625-7738	Sequence 7738, Ap
20	977.5	36.5	472	17	US-09-306-844-4	Sequence 4, Appl1
21	946	35.3	444	19	US-09-513-996A-55848	Sequence 55848, A
22	946	35.3	444	19	US-09-513-996A-56470	Sequence 56470, A
23	946	35.3	444	23	US-09-935-625-5228	Sequence 5228, Ap
24	946	35.3	444	23	US-09-935-625-5736	Sequence 5736, Ap
25	946	35.3	444	23	US-09-935-625-7740	Sequence 7740, Ap
26	946	35.3	444	23	US-09-935-625-9350	Sequence 9350, Ap
27	866	32.3	401	23	US-09-935-625-5737	Sequence 5737, Ap
28	866	32.3	401	23	US-09-935-625-9351	Sequence 9351, Ap
29	860	32.1	457	21	US-09-708-427-20715	Sequence 20715, A
30	832	31.0	499	26	US-10-319-999-40933	Sequence 40933, A
31	832	31.0	503	27	US-10-324-109-17051	Sequence 17051, A
32	832	31.0	526	27	US-60-337-358-384	Sequence 384, App
33	822	30.7	362	23	US-09-995-917A-1	Sequence 17051, A
34	816.5	30.5	486	25	US-10-142-231-92	Sequence 92, Appl1
35	808	30.1	427	21	PCT-US99-2285B-1487	Sequence 20716, A
36	805	30.0	153	1	PCT-US99-2285B-1487	Sequence 3487, Ap
37	792	29.5	518	27	US-10-219-999-43776	Sequence 43776, A
38	792	29.5	552	26	US-60-324-109-26814	Sequence 26814, A
39	792	29.5	552	27	US-10-219-999-49629	Sequence 49629, A
40	790	28.5	515	26	US-10-219-999-49629	Sequence 17918, A
41	761.5	28.4	464	21	US-09-791-537-17918	Sequence 381, App
42	754	28.1	489	27	US-60-337-358-381	Sequence 62903, A
43	754	28.1	490	26	US-10-219-999-62903	Sequence 31870, A
44	754	28.1	490	27	US-60-324-109-31870	Sequence 54439, A
45	739.5	27.6	497	26	US-10-219-999-54439	

ALIGNMENTS

```
RESULT 1
PCT-US00-03820-2
; Sequence 2, Application PC/TUS0003820
; GENERAL INFORMATION:
; APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA
; TITLE OF INVENTION: DWE4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001.40
; CURRENT APPLICATION NUMBER: PCT/US00/03820
; EARLIER FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: 60/119, 657
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/119, 658
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
PCT-US00-03820-2
```

```
Query Match 100.0%; Score 2681; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 3.4e-231;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFEHTHTLPLLLPSLLSLLFLLLKRRNRKTRNPPGSGMPFLGTTGYLKPVT 60
Db 1 MFEHTHTLPLLLPSLLSLLFLLLKRRNRKTRNPPGSGMPFLGTTGYLKPVT 60
```

QY	61	ATTIGDEMOOHVSKXGKIYRSMLPEEPTIVSADAGINRIITLONEGRLEFECSPRSIGILL	120
Db	61	ATTIGDEPMOQOHVSKGKIYRSMLPEEPTIVSADAGINRIITLONEGRLEFECSPRSIGILL	120
QY	121	GKWSMLVYVGDMHNRDMRSISLNFLESHARLRTLLDKDVERHTLFVLDSMOONSIFSAODEA	180
Db	121	GKWSMLVYVGDMHNRDMRSISLNFLESHARLRTLLDKDVERHTLFVLDSMOONSIFSAODEA	180
QY	181	KKFTTNLMAKHIHMSMDPGEEETEOULKEVYTFMKGVSAPLMPCGYAKHALOSRATYILK	240
Db	181	KKFTTNLMAKHIHMSMDPGEEETEOULKEVYTFMKGVSAPLMPCGYAKHALOSRATYILK	240
QY	241	FIERKMEERKLDIKEDEDEEBEEVKTEDDEAKMSKSDHVQRQRDDDLGAVLKHNSNSTEO	300
Db	241	FIERKMEERKLDIKEDEDEEBEEVKTEDDEAKMSKSDHVQRQRDDDLGAVLKHNSNSTEO	300
QY	301	ILDLILSLFLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKELGESLNMDD	360
Db	301	ILDLILSLFLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKELGESLNMDD	360
QY	361	YKKMDFQOCVINEITRLGNNVAREFLRKALKDVRKYGYDLPSCMKVLPVYSAVHLNDSRD	420
Db	361	YKKMDFQOCVINEITRLGNNVAREFLRKALKDVRKYGYDLPSCMKVLPVYSAVHLNDSRD	420
QY	421	QPNLEFNPMRMOQOONNGASSSGSGSPSTWGNNTNMPRGGGFRLCAGSELAKLEMAVFIHTHY	480
Db	421	QPNLEFNPMRMOQOONNGASSSGSGSPSTWGNNTNMPRGGGFRLCAGSELAKLEMAVFIHTHY	480
QY	481	LKFNNWELAEDDQPFAPFPVDFPNGLPPIRYSRTL	513
Db	481	LKFNNWELAEDDQPFAPFPVDFPNGLPPIRYSRTL	513

RESULT 2

Sequence 2 Application US/09502426
 GENERAL INFORMATION:
 APPLICANT: AZPIROZ, Ricardo
 APPLICANT: CHOE, Sungwa
 APPLICANT: FELDMANN, Kenneth
 TITLE OF INVENTION: DMFA POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
 FILE REFERENCE: 2225-0001
 CURRENT APPLICATION NUMBER: US/09/502,426
 CURRENT FILING DATE: 2000-02-11
 EARLIER APPLICATION NUMBER: 60/119,657
 EARLIER FILING DATE: 1999-02-11
 EARLIER APPLICATION NUMBER: 60/119,658
 EARLIER FILING DATE: 1999-02-11
 NUMBER OF SEQ. ID NOS.: 18
 SOFTWARE: PatentIn Ver. 2.0
 SEQ. ID NO. 2
 LENGTH: 513
 TYPE: PRT
 ORGANISM: Arabidopsis sp.
 US-09-502-426-2

Query Match	100.0%;	Score 2681;	DB 19;	Length 513;
Best Local Similarity	100.0%;	Pred. No. 3.4e-231;		
Matches 513;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MEFEHNHLPLLLPLSLSTLLFLILKRRNRKRRENLPPGKSGMPFGETIGYLKPYT	60
Db	1	MEFEHNHLPLLLPLSLSTLLFLILKRRNRKRRENLPPGKSGMPFGETIGYLKPYT	60
QY	61	ATTGADFMQOAVSKGYIKRSNLFGEPTIVSADAGINREITLONEGRLEFCSYPRISGGIL	120
Db	61	ATTGADFMQOAVSKGYIKRSNLFGEPTIVSADAGINREITLONEGRLEFCSYPRISGGIL	120
QY	121	GKMSMLVVGGMHRMDRSISLNFSLSHARLTTLTKDVEHNTLFVLDSMOONSIIFSQADEA	180
Db	121	GKMSMLVVGGMHRMDRSISLNFSLSHARLTTLTKDVEHNTLFVLDSMOONSIIFSQADEA	180
QY	181	KKPTFNIMAKHIMSDPGEETEOCLKKEVYTFMKGVASAPLNLPGTAVERALQSRATILK	240

Db	181	KFFETNLMAKHTMSMDPGEETETOLKKEVYTFPAKGYVSPALNPLGTAHYKALOSRATILK	240
QY	241	FIERKMEERKLDIKEEDOEEEVEKTEDEAEMSKSDHYRRQRTDDLLGWLKHSNLTSEQ	300
Db	241	FIERKMEERKLDIKEEDOEEEVEKTEDEAEMSKSDHYRRQRTDDLLGWLKHSNLTSEQ	300
QY	301	IIDLILSLFAGHETSSVAIALAIFLOACPAVEELREHETIARAKKELGSELSNWD	360
Db	301	IIDLILSLFAGHETSSVAIALAIFLOACPAVEELREHETIARAKKELGSELSNWD	360
QY	361	YKMDFTOCVINETTRLGNVVRFLHKKALKDVKYKDYDIPSGMKVLPVISAHLDSRYD	420
Db	361	YKMDFTOCVINETTRLGNVVRFLHKKALKDVKYKDYDIPSGMKVLPVISAHLDSRYD	420
QY	421	QPNLEPNPMWQOONNGASSSGSSTFGWGNNTYPPFGGGRILCAGSELAKIEMAVFIHLV	480
Db	421	QPNLEPNPMWQOONNGASSSGSSTFGWGNNTYPPFGGGRILCAGSELAKIEMAVFIHLV	480
QY	481	LKENNELAEDDOPFAFPFVDFPNGLPIRYSRLI	513
Db	481	LKENNELAEDDOPFAFPFVDFPNGLPIRYSRLI	513

RESULT 3

```

1  Sequence 67246, Application US/09791537
2  GENERAL INFORMATION:
3  APPLICANT: Biomonix, Inc.
4  APPLICANT: Debe, Derek
5  TITLE OF INVENTION: THREE DIMENSIONAL STR
6  TITLE OF INVENTION: METHODS OF USE THERE
7  FILE REFERENCE: 261/210
8  CURRENT APPLICATION NUMBER: US/09/791,537
9  CURRENT FILING DATE: 2001-02-22
10 NUMBER OF SEQ ID NOS: 153055
11 SOFTWARE: PatentIn version 3.0
12 SEQ ID NO 67246
13 LENGTH: 513
14 TYPE: PRT
15 ORGANISM: Arabidopsis thaliana
16 US-09-791-537-67246

```

Query Match	99.9%	Score 2677;	DB 21;	Length 513;
Best Local Similarity	99.8%	Pred. No. 7.8e-231;		
Matches 512; Conservative	1;	Mismatches .0;	Indels 0;	Gaps 0;

Qy	1	MEETHHTLLPLLLPLSLLSLLLELLLRANKKTFEPLPGKSGMPGETTGYLKPPT	60
Dd	1	MEETHHTLLPLLLPLSLLSLLLELLLRANKKTFEPLPGKSGMPGETTGYLKPPT	60
Qy	61	ATTLEDPMOAHYSKYK IYRSNLFGEPTTVSADAGNRFLLQNEGRLLFECYSRSIGIL	120
Dd	61	ATTLEDPMOAHYSKYK IYRSNLFGEPTTVSADAGNRFLLQNEGRLLFECYSRSIGIL	120
Qy	121	GKMSNLVIVGDHNRDMRSISLNFLSHARLRTLLKDYERHTTVLDSMOONSIFSAODEA	180
Dd	121	GKMSNLVIVGDHNRDMRSISLNFLSHARLRTLLKDYERHTTVLDSMOONSIFSAODEA	180
Qy	181	KKFTFNLMAKHIMSDPGEETPEOLKKEYVTMKGVASAPLNLPGTAYHKALOSRAATILK	240
Dd	181	KKFTFNLMAKHIMSDPGEETPEOLKKEYVTMKGVASAPLNLPGTAYHKALOSRAATILK	240
Qy	241	FIERKMERKLDIKKEEDOEVEEVKTKEDAEEMSKSDHYVRQRDDDLLGWLLKXSNLSTEG	300
Dd	241	FIERKMERKLDIKKEEDOEVEEVKTKEDAEEMSKSDHYVRQRDDDLLGWLLKXSNLSTEG	300
Qy	301	ILDILISLLPFGHEHSSVAIALATIFPLQCPAAYELREHNEIARAKKELGESLNMWD	360
Dd	301	ILDILISLLPFGHEHSSVAIALATIFPLQCPAAYELREHNEIARAKKELGESLNMWD	360
Qy	361	YKKMDFTCVINEITRLGNVAVFLHRRKALKDVRKYGYDIPSGMKVLPVIVSAVLLDMSRYD	420

```

Db      361 YKKMDFQCYNINELRLGNVRFHRRALKDVKRYKGYDIPSGMKVLPVISAHLNDRYD 420
QY      421 QPNLFNPMRMOQNNNGSSSGSFSFTWGNVMPFGGGRPCACSELAKLEMAVFIHLLV 480
Db      421 QPNLFNPMRMOQNNNGSSSGSFSFTWGNVMPFGGGRPCACSELAKLEMAVFIHLLV 480
QY      481 LKFNWELAEDDQPAFPFVDFPENGPIRVSRI 513
Db      481 LKFNWELAEDDQPAFPFVDFPENGPIRVSRI 513

```

RESULT 4

```

US-09-464-535-36
; Sequence 36, Application US/09464535
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Cahoou, Rebecca E.
; APPLICANT: Sakai, Hajime
; APPLICANT: Mcgonigle, Brian
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: STEROL BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB1306 US NA
; CURRENT APPLICATION NUMBER: US/09/464,535
; EARLIER FILING DATE: 1999-12-15
; EARLIER APPLICATION NUMBER: 60/112,555
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-464-535-36

```

```

Query Match      55.9%; Score 1500; DB 18; Length 392;
Best Local Similarity 70.3%; Pred. No. 2.9e-125;
Matches 289; Conservative 40; Mismatches 54; Indels 28; Gaps 2;

```

```

QY      102 QNEGRLEPCSTPRSTIGGILGKWSMLVGVGMHGRMRSISLNFSLHARKLTLLKDVREHT 161
Db      3 QNEGRLEPCSTPRSTIGGILGKWSMLVGVGMHGRMRSISLNFSLHARKLTLLKDVREHT 62
QY      162 LFFVDSMOQNSIFEAODEAKKFTFNLMAKHNSMDPGEETEOLKKEVTFMKGVSAVL 221
Db      63 LFLVRAWPPSTFSQHOAKKFTFNLMAKHNSMDPGEETEOLKKEVTFMKGVSAVL 122
QY      222 NLPGTAYHAKALQSRAITLFTERKMEERKIDKEEDOEDEEVKTEDEAEMKSDHVRKOR 281
Db      123 NLPGTAYHAKALQSRAITLFTERKMEERKIDKEEDOEDEEVKTEDEAEMKSDHVRKOR 163
QY      282 TDDDLGKVLKHSNSTQIIDLILSLFAGHETSSVAIALAIFPLQACPKVAEELREH 341
Db      164 --DDLGLAKLQNSISKEQIIDLILSLFAGHETSSVAIALAIFPLQACPKVAEELREH 221
QY      342 LEIARAKKEGESELENDYKMDTQCVINETRLGNVRFHRRALKDVKRYKGYDIPS 401
Db      222 LGIARQRLRECKLSMEDYKEMFTQCVINETRLGNVRFHRRALKDVKRYKGYDIPS 281
QY      402 GKMVLPIVISAHLNDRYDQPNLFNPMRMOQNNNGSSSGSFSFTWGNVMPFGGGRPC 461
Db      282 GKMVLPIVISAHLNDRYDQPNLFNPMRMOQNNNGSSSGSFSFTWGNVMPFGGGRPC 334
QY      462 CAGSELAKLEMAVFIHLLVLFKFNWELAEDDQPAFPFVDFPENGPIRVSRI 512
Db      335 CAGSELAKLEMAVFIHLLVLFKFNWELAEDDQPAFPFVDFPENGPIRVSRI 385

```

RESULT 5

```

US-09-708-427-51512
; Sequence 51512, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.

```

```

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51512
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..484
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..484
; OTHER INFORMATION: Ceres Seq. ID 1927049
US-09-708-427-51512

```

```

Query Match      37.9%; Score 1016; DB 21; Length 484;
Best Local Similarity 41.4%; Pred. No. 1.2e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY      7 HTLLPL--LLPLSLSL--FLILRRNRKTFNLPKSGMPFGETIGYLPKPYAT 62
Db      9 HXLAALCTLLASLTLLVNLHFLPL--LNPKA----PRSGFGLGETLRLFTPLASN 62
QY      63 TLGDFMOQHVSKYKIRSNLFGPTTVSADAGINRFTLONEGRLEPCSTPRSTIGILGK 122
Db      63 TLGDFLHDHCSRYGRVFKSHLFCPTVAVSCDDQDLNHFILONEGRLEPCSTPRSTIGILGK 122
QY      123 WSMVLVGVGMHGRMRSISLNFSLHARKLTLLKDVREHTLFFVDSMOQNS--IFSA 176
Db      123 SSMVLVGVGMHGRMRSISLNFSLHARKLTLLKDVREHTLFFVDSMOQNS--IFSA 182
QY      177 QDEAKKFTFNLMAKHNSMDPGEETEOLKKEVTFMKGVSAVLNPGTAYHAKALQSR 236
Db      183 CEARKFAFSYVQVGLSPEEYVTAILEDPLAFKGLSPFLYIPGTPYAKAVARE 242
QY      237 TLKFERKMEERKIDKEEDOEDEEVKTEDEAEMKSDHVRKORTDDLLGVNLSNL 296
Db      243 RISSTVKGILKER-----SAGSMNKQ---GDFLVDLSSNEL 277
QY      297 STQIIDLILSLFAGHETSSVAIALAIFPLQACPKVAEELREHLEIARAKKEGESE 355
Db      278 SDEKVSFVLDLILGGETYSLLISNVVFLGQSAQDLQVLRHDSI--RSNK--GKEEC 334
QY      356 LNMDDYKMDTQCVINETRLGNVRFHRRALKDVKRYKGYDIPSGMKVLPVISAHLN 415
Db      335 LNSDYKMEYTOQVINEALRCGNIVAFVHRRALKDVKRYKGYDIPSGMKVLPVISAHLN 394
QY      416 NSRYDQPNLFNPMRMOQNNNGSSSGSFSFTWGNVMPFGGGRPCACSELAKLEMAV 475
Db      395 PSLGDXQGFQPCRMESTSGTS-----KRPTEPGGGRPLXPSSELAVETAF 443
QY      476 LHHVLKFNWELAEDDQPAFPFVDFPENGPIRVSRI 509
Db      444 LHHVLKFNWELAEDDQPAFPFVDFPENGPIRVSRI 477

```

RESULT 6

```

US-09-708-427-51511
; Sequence 51511, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2250-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 51511
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..485
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..485
; OTHER INFORMATION: Ceres Seq. ID 1927048
US-09-708-427-51511

```

```

Query Match          37.9%; Score 1016; DB 21; Length 485;
Best Local Similarity 41.4%; Pred. No. 1.2e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY 7 HTLLPL--LLPSLLSL--FLILKRRNRKTRFNPGRKSGMPTGEGTIGLKRYAT 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 HXLAALGTLTLLASLTLLVNLHFLPL--LNPKA---PRGSFGWPLGELTRLPHTASN 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 TLGDPMOQHVSKYKIRSNLGEPTIVSADAGLNRFILQNGRLFECSYPSRIGTIGK 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 TLGGLFDHCSRYGVFKSHLFCPTIVSCDQDLNHLFQNGERLFGCSYPRPHIGILGK 123
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 WSMVLVGDMDHRMDSISLNFSLHARLRTILKDYERHTLVLDSSQONS-----IFSA 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 SSMVLVVGEDHRRNLALALVSTKLKPSYLGDIKIALHYVGAMRRHSSGCVRYAVAF 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 ODEAKKFTFNMAKHIMSDPGEETEOQKREYTFMKGVSAPLNLPGTAVHKAQSORA 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 184 CEARKKFAFSYIVKOVGLSPPEPVARTLEDFLAFMKGLISFPLTPTGPAKAVRARE 243
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 TLKFIERKMEERKLDIKEEDQEEBEVKTEDEAEMSKSDHYRKORTDDLLGVNLKHSNL 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 244 RISSTVKGILIKRR-----SAGSNMKQ---GDFLDVLLSSNEL 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 STEQILDLLSLFLAGHETSSVALAIFLQACPKAVEELREHLEIARAKKELGESE- 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 279 SDEKVSFVLDLISLGGYETSTLISMVYVFLQSOADDLVLRREHDSI--RSNK--GKEEC 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 356 LNMDDYKMDPQOCINETRLGNVRFHRRKALDVRKGYDIPSGKVLVPLVSAVHD 415
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 LISEDYKMEYTOQVINEALRCGNIVKFAHRRKALDKVKEYLLPSGKVLVPLVFAVHLN 395
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 416 NSRYDQPNLFNPMWMOQONNGASSGGSFSTWGNMNYMPFGGPRLCAGSELAKLEMAVF 475
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 PSLHGDXXQFOPCRMEGTSQGS-----KRFPPFGGPRLXGSELAKVETAF 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 476 IHHVLKFNWELAEDDQPPAFPFVDFPNCGLPIRV 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 LHHLLXNLRWRKIXDDIPMAYPYXXFQRLPIEI 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 7

```

US-09-708-427-51510
; Sequence 51510, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708, 427
; EARLIER FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51510
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..542

```

```

; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..542
; OTHER INFORMATION: Ceres Seq. ID 1927047
US-09-708-427-51510

```

```

Query Match          37.9%; Score 1016; DB 21; Length 542;
Best Local Similarity 41.4%; Pred. No. 1.4e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY 7 HTLLPL--LLPSLLSL--FLILKRRNRKTRFNPGRKSGMPTGEGTIGLKRYAT 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 HXLAALGTLTLLASLTLLVNLHFLPL--LNPKA---PRGSFGWPLGELTRLPHTASN 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 TLGDPMOQHVSKYKIRSNLGEPTIVSADAGLNRFILQNGRLFECSYPSRIGTIGK 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TLGGLFDHCSRYGVFKSHLFCPTIVSCDQDLNHLFQNGERLFGCSYPRPHIGILGK 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 WSMVLVGDMDHRMDSISLNFSLHARLRTILKDYERHTLVLDSSQONS-----IFSA 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 SSMVLVVGEDHRRNLALALVSTKLKPSYLGDIKIALHYVGAMRRHSSGCVRYAVAF 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 ODEAKKFTFNMAKHIMSDPGEETEOQKREYTFMKGVSAPLNLPGTAVHKAQSORA 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 CEARKKFAFSYIVKOVGLSPPEPVARTLEDFLAFMKGLISFPLTPTGPAKAVRARE 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 TLKFIERKMEERKLDIKEEDQEEBEVKTEDEAEMSKSDHYRKORTDDLLGVNLKHSNL 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 RISSTVKGILIKRR-----SAGSNMKQ---GDFLDVLLSSNEL 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 STEQILDLLSLFLAGHETSSVALAIFLQACPKAVEELREHLEIARAKKELGESE- 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 SDEKVSFVLDLISLGGYETSTLISMVYVFLQSOADDLVLRREHDSI--RSNK--GKEEC 392
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 356 LNMDDYKMDPQOCINETRLGNVRFHRRKALDVRKGYDIPSGKVLVPLVSAVHD 415
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 393 LISEDYKMEYTOQVINEALRCGNIVKFAHRRKALDKVKEYLLPSGKVLVPLVFAVHLN 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 416 NSRYDQPNLFNPMWMOQONNGASSGGSFSTWGNMNYMPFGGPRLCAGSELAKLEMAVF 475
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 PSLHGDXXQFOPCRMEGTSQGS-----KRFPPFGGPRLXGSELAKVETAF 501
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 476 IHHVLKFNWELAEDDQPPAFPFVDFPNCGLPIRV 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 LHHLLXNLRWRKIXDDIPMAYPYXXFQRLPIEI 535
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8

```

US-09-306-844-2
; Sequence 2, Application US/09306844
; GENERAL INFORMATION:
; APPLICANT: Konec, Jasda
; APPLICANT: Mathur, Jaldeep
; APPLICANT: Szekeres, Miklos
; APPLICANT: Altman, Thomas
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME P450-TYPE
; TITLE OF INVENTION: PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS IN
; FILE REFERENCE: 147-186
; CURRENT APPLICATION NUMBER: US/09/306, 844
; EARLIER FILING DATE: 1999-05-06
; EARLIER FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-306-844-2

```

```

Query Match          36.5%; Score 978.5; DB 17; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.7e-78;

```

```

OY      132  MHRMRISINFSHARLFTLLKQVERTLLVDSWOONSIFSAQDAAKKEFTNMAKH 191
Db      124  LHRKMSLJMSFANSSIKDHLMDLIDRLVFRNNLDSMSRYLL--MEBAKITEFLVQO 181
OY      192  IMSMDPEEETDLKREYVTFPMKGVSAVLNPGTAUKALQSRITLIKFERKMEERKL 251
Db      182  IMSDPDQ--EMSESLRKEVLLIEGFSULPLIFSTYTKAIQAR-----RKVALALT 232
OY      252  DIKEEDDEEEYVTEDEBAEMSKSDHVKQRTDDLLGVNLKHSNLSTEQILDLILSLFA 311
Db      233  VVVMKREEEBEEAE-----RKQMDLALLAADDGSPDEIYDFLVALIVA 278
OY      312  GHETSVALALATEFLQACPAVEELREBHEILIAKKELGESLMDQVKKMDFTQCVI 371
Db      279  GYETTSITMLAKVFLTEPLALADQKEHEKIRAMKSD--SYSLMSDYKSMPTQCVV 336
OY      372  NETLRLGNVRLHRRKALDVRVYKGDIPSGKVLVYTSVAHLNDSRPDQNPENPRMQ 431
Db      337  NETLRAANIIGVFRRAMTDVEILKQKPIKPKGKAVSSSRAYHLDPNHFKDARFTNPRMQ 396
OY      432  QQNNNGASSGSGSGSTWGN--YMPGGGPRICAGSELAKLEMAVFIHLVLKFNWELAE 490
Db      397  -----SNSVTTGPSNVFTPPGGGRILCPGYELARVALSVFLHRLVTGGSWPABQ 446
OY      491  DQFAFPYDFPNGLPIRYSR 511
Db      447  DKLVPEPRTTRQKRYPIFYKR 467

```

[illegible]

QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 233 VVVKRRREEEEGAE-----RKMDLAALLAADGFSDEIDVDFVALLVA 278
QY 312 GHETSSVAIALAIFFLQACRAVEELREHELEIARAKKELGESLNMDDYKMDTQCVI 371
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 279 GYETTSITMTLAVFLETETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNNVFLHRLKALDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNPMQ 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 337 NETLRVANIIGVFRRAATDVEIKGYKIPKGMKVFSSRAVHLDPNHFKAARTFPMRMQ 396
QY 432 QONNGASSSGSSTWGN--YMPGGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 397 -----SNSVTTGPSNVFTPPGGGPRLCPEGLARVALSVFLHRLVTFGFSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 447 DKLVFFPTTRQKRPPIFVKR 467

RESULT 11
US-09-791-537-145750
; Sequence 145750, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Dede, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 145750
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-791-537-145750

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 2,7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLILFLILKRRNKTRENLPGKSGMPLETIGYIKPYATTLGDFMOOH 71
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 7 LLLSSIAAGFL--LLRRTYRRMGILPGSLGLPIGETFOLIGATKTEPPEPTIDER 63
QY 72 VSKYKIRYRNLFGEPITVSADAGLNRFILONEGRLEFCSYPRSIGILGKMSMLVYGD 131
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 64 VARGSVFMTHLFGEPITVSADPETNRFLVONEGKLEFCSYPASICNLLGKSLLMKGS 123
QY 132 MHRDMRSISLNFSLHARLITLKDVERHTLFLVDSMOONSIFSAODEAKKTFENLMK 191
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 124 LHKRMSLJMSFANSIHKDMLDIDRLVRFNDSMSRYLL--MEBAKKTTFELVYQ 181
QY 192 IMSNDPGEETEOLKKEVTFMGVVSAPLNPCTAYHKAQSRATILKFERKMEERKL 251
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 182 LMSFDPG--EWSESLRKEILVIEGFSILPLPFLSTTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 233 VVVKRRREEEEGAE-----RKMDLAALLAADGFSDEIDVDFVALLVA 278
QY 312 GHETSSVAIALAIFFLQACRAVEELREHELEIARAKKELGESLNMDDYKMDTQCVI 371
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 279 GYETTSITMTLAVFLETETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNNVFLHRLKALDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNPMQ 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 337 NETLRVANIIGVFRRAATDVEIKGYKIPKGMKVFSSRAVHLDPNHFKAARTFPMRMQ 396

QY 432 QONNGASSSGSSTWGN--YMPGGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 397 -----SNSVTTGPSNVFTPPGGGPRLCPEGLARVALSVFLHRLVTFGFSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 447 DKLVFFPTTRQKRPPIFVKR 467

RESULT 12
US-09-935-625-5227
; Sequence 5227, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CA
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 5227
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 2114147
US-09-935-625-5227

Query Match 36.5%; Score 978.5; DB 23; Length 472;
Best Local Similarity 41.9%; Pred. No. 2,7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLILFLILKRRNKTRENLPGKSGMPLETIGYIKPYATTLGDFMOOH 71
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 7 LLLSSIAAGFL--LLRRTYRRMGILPGSLGLPIGETFOLIGATKTEPPEPTIDER 63
QY 72 VSKYKIRYRNLFGEPITVSADAGLNRFILONEGRLEFCSYPRSIGILGKMSMLVYGD 131
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 64 VARGSVFMTHLFGEPITVSADPETNRFLVONEGKLEFCSYPASICNLLGKSLLMKGS 123
QY 132 MHRDMRSISLNFSLHARLITLKDVERHTLFLVDSMOONSIFSAODEAKKTFENLMK 191
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 124 LHKRMSLJMSFANSIHKDMLDIDRLVRFNDSMSRYLL--MEBAKKTTFELVYQ 181
QY 192 IMSNDPGEETEOLKKEVTFMGVVSAPLNPCTAYHKAQSRATILKFERKMEERKL 251
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 182 LMSFDPG--EWSESLRKEILVIEGFSILPLPFLSTTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 233 VVVKRRREEEEGAE-----RKMDLAALLAADGFSDEIDVDFVALLVA 278
QY 312 GHETSSVAIALAIFFLQACRAVEELREHELEIARAKKELGESLNMDDYKMDTQCVI 371
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 279 GYETTSITMTLAVFLETETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNNVFLHRLKALDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNPMQ 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 337 NETLRVANIIGVFRRAATDVEIKGYKIPKGMKVFSSRAVHLDPNHFKAARTFPMRMQ 396
QY 432 QONNGASSSGSSTWGN--YMPGGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 397 -----SNSVTTGPSNVFTPPGGGPRLCPEGLARVALSVFLHRLVTFGFSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 447 DKLVFFPTTRQKRPPIFVKR 467

RESULT 13
US-09-935-625-5735

; Sequence 5735, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 5735
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 3058391
US-09-935-625-5735

Query Match 36.5%; Score 978.5; DB 23; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLLSLFLILKRRNRKTRFNPGLPKSGMPFLGFTIGYIKPYTATLGDPMQOH 71
DB 7 LLLSSIAAGFL--LLLRTRRYRMGLPPLSGLPGLGFTFOLIGATKTEPPEFIDER 63
QY 72 VSKYKITYRSNLFCEPTIVSADAGLNRLITLONEGRLEFCSYPRISGIGILGKSMVLVYGD 131
DB 64 VARGSVPMTHLFREPTIFSDPTNRRVLONEGKLEFCSYPASICNLGKSHLLMGKS 123
QY 132 MHRDMRSISLNFSLHARLRTLLKDVRRHTLFLVDSMQONSIFSAODEAKKFTNLMK 191
DB 124 LHKRHSLSLMSFANSSTIKDHLMDIDLVRFNDSWSRYLL--MEAKKITFELTVKQ 181
QY 192 IMSDPEEETEOLKREYVTPMKGVASAPLNPCTAHHKALQSATIIKFERKMEERKL 251
DB 182 LMSFDPG-EMSESLRKEYLLVIEGFSLPLPLFSTYTKAIOAR-----RKVAEALT 232
QY 252 DIKEEDOEVEEYKTEDEAEKSKSDHVRKORTDDLLGVKLKHSNLSQIIDLILSLFA 311
DB 233 VVVMKRRREEEGAE-----RKMDLAALLAADDFSDDEIVDFVALVLA 278
QY 312 GHETSSVAIALAIFLOACPRAVEELREHELEIARAKEGESELMNDYKKMDPTQCVI 371
DB 279 GYETSTIMTLAVKFLTETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRIGNVRFLLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSRYDQPLFNPWMQ 431
DB 337 NETLRVANIIGVFRAMTVEIKGYKIPKGMKVFSSFRVAHLDPNHEKARTFNPWMQ 396
QY 432 QONNGASSGSGSFTWGN--YMPFGGPRLCAGSELAKLEMAVFIHHLVLEKFMWELAD 490
DB 397 -----SNVTTGPSNVFTFPGGPRLCPEYELARVALSVFLHRLVTFGSVWPAEQ 446
QY 491 DQPAPEVDVDPNGLPPIVRSR 511
DB 447 DKLVEFFPTTRTKRYPIFYKR 467

RESULT 14
; Sequence 7739, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 7739
; LENGTH: 472

; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 2114147
US-09-935-625-7739

Query Match 36.5%; Score 978.5; DB 23; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLLSLFLILKRRNRKTRFNPGLPKSGMPFLGFTIGYIKPYTATLGDPMQOH 71
DB 7 LLLSSIAAGFL--LLLRTRRYRMGLPPLSGLPGLGFTFOLIGATKTEPPEFIDER 63
QY 72 VSKYKITYRSNLFCEPTIVSADAGLNRLITLONEGRLEFCSYPRISGIGILGKSMVLVYGD 131
DB 64 VARGSVPMTHLFREPTIFSDPTNRRVLONEGKLEFCSYPASICNLGKSHLLMGKS 123
QY 132 MHRDMRSISLNFSLHARLRTLLKDVRRHTLFLVDSMQONSIFSAODEAKKFTNLMK 191
DB 124 LHKRHSLSLMSFANSSTIKDHLMDIDLVRFNDSWSRYLL--MEAKKITFELTVKQ 181
QY 192 IMSDPEEETEOLKREYVTPMKGVASAPLNPCTAHHKALQSATIIKFERKMEERKL 251
DB 182 LMSFDPG-EMSESLRKEYLLVIEGFSLPLPLFSTYTKAIOAR-----RKVAEALT 232
QY 252 DIKEEDOEVEEYKTEDEAEKSKSDHVRKORTDDLLGVKLKHSNLSQIIDLILSLFA 311
DB 233 VVVMKRRREEEGAE-----RKMDLAALLAADDFSDDEIVDFVALVLA 278
QY 312 GHETSSVAIALAIFLOACPRAVEELREHELEIARAKEGESELMNDYKKMDPTQCVI 371
DB 279 GYETSTIMTLAVKFLTETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRIGNVRFLLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSRYDQPLFNPWMQ 431
DB 337 NETLRVANIIGVFRAMTVEIKGYKIPKGMKVFSSFRVAHLDPNHEKARTFNPWMQ 396
QY 432 QONNGASSGSGSFTWGN--YMPFGGPRLCAGSELAKLEMAVFIHHLVLEKFMWELAD 490
DB 397 -----SNVTTGPSNVFTFPGGPRLCPEYELARVALSVFLHRLVTFGSVWPAEQ 446
QY 491 DQPAPEVDVDPNGLPPIVRSR 511
DB 447 DKLVEFFPTTRTKRYPIFYKR 467

RESULT 15
; Sequence 9349, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 9349
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 3058391
US-09-935-625-9349

Query Match 36.5%; Score 978.5; DB 23; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

0Y 12 LLLLSLLSLLELLFLLIKLRNKRKTRPNLPBGSGMFLGELTYGLKLPYATATLGDGMQOH 71
Db 7 LLLLSIAAGFL--LLLRTRYRRMGRLPBGSLGLPLIETPOLIAYATENPEPTIDER 63
0Y 72 VSKYKIXRSLNEGEPPTVSADAGINRFLONEGRLEFECSPRSTIGILLGKSMYLAVGD 131
Db 64 VARYSVFMTHLFGEPPTFVSADPENRNFVLONEGKLFECSSYPASICNLGKSHLLMKGK 123
0Y 132 MHRORSTISINFLSHARLTLLKQYERTLLVLSMOONSIFSAODEKKKTFNLMKX 191
Db 124 LKRRHSLTMSWANSIILKDHLLDLDRLVRENLDWSRSSVLL--MEKKKTFTELTVAQ 181
0Y 192 IMSMDPGEETEQOLKEVYTFMKGVYSAFLNPGTAHYHALOSRATILFIERKMEERKL 251
Db 182 LMSFPDGP--EMSESLRKREYLLVEGFPSLPLPLESTYYRAIQAR-----RKVAEALT 232
0Y 252 DIKEDEDEEEVKTDEDEAMSKSDHYRKRORTDDLLGWLKHNSLSTEQILDLILSLRA 311
Db 233 VVMKRRREEEGAE-----RKKMDLAALLAADGFSDEEIVPVALVLA 278
0Y 312 GHETSSVAIALAIFELQACPRAVEILREHEILARAKKELGESLEMMDDYKKKDFPQCVI 371
Db 279 GYETSTIMTLAVKFLTEPLPLAQLKDEHEKIRAKNSD--SYLEMSDYKSMPLFOCV 336
0Y 372 NETLALGVNVRLLHKKALKDVKYKGYDIPDSGKKVLPVISAYVHLDNSRTQOPLFNWRMO 431
Db 337 NETLEVNAMIGGVFRAMTDVEIKKIKPKGKKVSSFRVAVHLDPMHFKDARTFNWRMO 396
0Y 432 OONNGASSSGSSEFSTWGN--YMPGSGGRPLCAGSELALEMVAIYHNLVLFKNWEALAD 490
Db 397 -----SNSVTTPGSPNVFTPPGGGRPLCPGYELAVALVSLFHLRLVTGSSWPAID 446
0Y 491 DQFAFPVDPFNGPLPIKYSR 511
Db 447 DKLVEFPTRQOKRYPFVKR 467

```

Search completed: June 10, 2003, 22:02:53
Job time : 163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:58:58 ; Search time 58 Seconds

(Without alignments)
2004.717 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681

Sequence: 1 MFETHTLLPLLLPSLLS.....FAFPYDFPGLPIRVSRLI 513

Scoring table:

BL0SUM62
Gap 10.0, Gapext 0.5

Searched: 1050051 seqs, 226654218 residues

Total number of hits satisfying chosen parameters: 1050051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA.New.*
1: /cgn2_6/pdata/2/paa/PC7_NEW_COMB.pep.*
2: /cgn2_6/pdata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/pdata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/pdata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/pdata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/pdata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/pdata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	US-09-502-426a-2	Sequence 2, Appl
2	2681	100.0	513	US-09-502-426b-2	Sequence 2, Appl
3	1770	66.0	506	US-10-395-463-2	Sequence 2, Appl
4	1166.5	43.5	300	US-10-424-599-282744	Sequence 282744,
5	1164	43.4	317	US-10-425-114-72982	Sequence 72982, A
6	1001	37.3	488	US-10-424-599-178693	Sequence 178693,
7	978.5	36.5	472	US-09-502-426a-19	Sequence 19, Appl
8	978.5	36.5	472	US-09-502-426b-19	Sequence 19, Appl
9	856.5	31.9	501	US-10-395-463-24	Sequence 24, Appl
10	849	31.7	575	US-09-502-426a-25	Sequence 25, Appl
11	849	31.7	575	US-09-502-426b-25	Sequence 25, Appl
12	832	31.0	503	US-10-310-154-384	Sequence 384, App
13	816.5	30.5	486	US-10-356-153-92	Sequence 92, App
14	792	29.5	518	US-10-425-114-383	Sequence 383, App
15	790	29.5	515	US-10-425-114-383	Sequence 383, App
16	761.5	28.4	464	US-09-502-426a-20	Sequence 20, Appl
17	761.5	28.4	464	US-09-502-426b-20	Sequence 20, Appl
18	754	28.1	489	US-10-310-154-381	Sequence 381, App
19	748	27.9	469	US-10-395-463-30	Sequence 30, Appl
20	739.5	27.6	497	US-10-425-114-64364	Sequence 64364, A
21	719.5	26.8	504	US-10-424-599-286550	Sequence 286550,
22	642	23.9	453	US-10-259-194A-156	Sequence 156, App
23	641	23.9	294	US-10-425-114-48987	Sequence 48987, A
24	634	23.6	470	US-10-310-154-386	Sequence 386, App
25	622	23.2	495	US-10-425-114-47304	Sequence 47304, A
26	614.5	22.9	496	US-10-424-599-271221	Sequence 271221,

27	602.5	22.5	499	US-10-310-154-382	Sequence 382, App
28	597.5	22.3	485	US-10-356-153-88	Sequence 88, Appl
29	596.5	22.2	497	US-10-356-153-64	Sequence 64, Appl
30	596	22.2	474	US-10-424-599-275182	Sequence 275182,
31	592	22.1	496	US-10-356-153-90	Sequence 90, Appl
32	588	21.9	509	US-10-356-153-67	Sequence 67, Appl
33	585	21.8	507	US-10-356-153-65	Sequence 65, Appl
34	583	21.7	484	US-10-356-153-57	Sequence 57, Appl
35	571	21.3	501	US-10-356-153-58	Sequence 58, Appl
36	569.5	21.2	493	US-10-356-153-89	Sequence 89, Appl
37	568.5	21.2	474	US-10-424-599-248036	Sequence 248036,
38	568.5	21.2	500	US-10-356-153-59	Sequence 59, Appl
39	564	21.0	484	US-10-356-153-56	Sequence 56, Appl
40	564	21.0	500	US-10-356-153-61	Sequence 61, Appl
41	559	20.9	298	US-10-424-599-205981	Sequence 205981,
42	558.5	20.8	483	US-10-356-153-62	Sequence 62, Appl
43	558.5	20.8	512	US-10-356-153-87	Sequence 87, Appl
44	553.5	20.6	498	US-10-356-153-91	Sequence 91, Appl
45	545	20.3	280	US-10-425-114-57231	Sequence 57231, A

ALIGNMENTS

RESULT 1
US-09-502-426a-2
; Sequence 2, Application US/09502426A
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHO, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DWf POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426A
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-502-426a-2

Query Match 100.0%; Score 2681; DB 5; Length 513;
Best Local Similarity 100.0%; Pred. No. 8.2e-25;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MFETHTLLPLLLPSLLSLLFTLLKRRNRKTRFNLPPKSGMPFLGETTGYLKYPT 60
1 MFETHTLLPLLLPSLLSLLFTLLKRRNRKTRFNLPPKSGMPFLGETTGYLKYPT 60
61 ATTLDPMQOHVSKTKYRSMLEPPTVSDAGLNRIILONEGRLEECSPRIGTIL 120
61 ATTLDPMQOHVSKTKYRSMLEPPTVSDAGLNRIILONEGRLEECSPRIGTIL 120
121 GKMSMLVYVGDHNRDMSISLNFSLSHARLRTLLKDVRRHTFLVDSMOONSIFSAODEA 180
121 GKMSMLVYVGDHNRDMSISLNFSLSHARLRTLLKDVRRHTFLVDSMOONSIFSAODEA 180
181 KFTFNMAKHTMSDPEEETBOLKKEVTFMKGVSAPIPLPTAVHKAQSRATILK 240
181 KFTFNMAKHTMSDPEEETBOLKKEVTFMKGVSAPIPLPTAVHKAQSRATILK 240
241 FIERMERKRDIDKEEDDEEVEKTEDEAKSKSDHVKORDDDLGWLKHSNLSTEQ 300
241 FIERMERKRDIDKEEDDEEVEKTEDEAKSKSDHVKORDDDLGWLKHSNLSTEQ 300
301 ILDLISLLFAGHTSSVAIALATFFLOACRAVEELREHLEIARAKKEGSELNMD 360
|||||

Db 301 IIDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELNWD 360
QY 361 YKAMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKAMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIHHLV 480
QY 481 LKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 513
Db 481 LKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 513

RESULT 2
US-09-502-426b-2
; Sequence 2, Application US/09502426B
; GENERAL INFORMATION:
; APPLICANT: Azpiloz, Ricardo
; APPLICANT: Choe, Sunghea
; APPLICANT: Feldman, Kenneth A.
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 11696-070001
; CURRENT APPLICATION NUMBER: US/09/502,426B
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: US 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 513
; TYPE: PRY
; ORGANISM: Arabidopsis thaliana
US-09-502-426b-2

Query Match 100.0%; Score 2681; DB 5; Length 513;
Best Local Similarity 100.0%; Pred. No. 8,2e-252;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFTEHNTLLPLLLPLSLLSLLFLILKRNKTRNLPFGKSGMFLGETTGYLKPYT 60
Db 1 MFTEHNTLLPLLLPLSLLSLLFLILKRNKTRNLPFGKSGMFLGETTGYLKPYT 60
QY 61 ATLDGPMQOHVSKGYIYSNLFGEPTIYSADAGLRFTILONGRLEFECSTPRSIGIL 120
Db 61 ATLDGPMQOHVSKGYIYSNLFGEPTIYSADAGLRFTILONGRLEFECSTPRSIGIL 120
QY 121 GKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAODEA 180
Db 121 GKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAODEA 180
QY 181 KRFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAITLK 240
Db 181 KRFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAITLK 240
QY 241 FIERKMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNLSTEQ 300
Db 241 FIERKMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNLSTEQ 300
QY 301 IIDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELNWD 360
Db 301 IIDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELNWD 360
QY 361 YKAMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKAMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIHHLV 480

QY 481 LKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 513
Db 481 LKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 513

RESULT 3
US-10-395-463-2
; Sequence 2, Application US/10395463
; GENERAL INFORMATION:
; APPLICANT: Tanaka, Hiroshi
; APPLICANT: Matsuo, Toshaki
; APPLICANT: Matsuo, Makoto
; APPLICANT: Iwahori, Shuichi
; TITLE OF INVENTION: Method of Controlling Character of Monocotyledon by Modified
; TITLE OF INVENTION: Overexpression of Cytochrome P450 Monooxygenase Gene Invol-
; TITLE OF INVENTION: Biosynthesis and Monocotyledon Modified by the Gene
; FILE REFERENCE: 59150-8022
; CURRENT APPLICATION NUMBER: US/10/395,463
; CURRENT FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: JP 2002-276398
; PRIOR FILING DATE: 2002-09-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 506
; TYPE: PRY
; ORGANISM: Oryza sativa
US-10-395-463-2

Query Match 66.0%; Score 1770; DB 6; Length 506;
Best Local Similarity 66.6%; Pred. No. 3.5e-163;
Matches 343; Conservative 56; Mismatches 74; Indels 42; Gaps 4;

QY 12 LLLLP-SLLSLPLILKRR-----NRKRFNLPFGKSGMFLGETTGYLK 57
Db 12 LLLLP-SLLSLPLILKRR-----NRKRFNLPFGKSGMFLGETTGYLK 57
QY 58 PYTATTLDDFMQOHVSKGYIYSNLFGEPTIYSADAGLRFTILONGRLEFECSTPRSIG 117
Db 58 PYTATTLDDFMQOHVSKGYIYSNLFGEPTIYSADAGLRFTILONGRLEFECSTPRSIG 117
QY 73 AHPATSVRFMEQHARCKIYRSLSFERVYVSADALNRTILONGRLEFECSTPRSIG 132
Db 73 AHPATSVRFMEQHARCKIYRSLSFERVYVSADALNRTILONGRLEFECSTPRSIG 132
QY 118 GILKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAQ 177
Db 118 GILKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAQ 177
QY 133 GILKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAQ 192
Db 133 GILKWSMLVVGDMHRMRSISLNFLSHARLRTILKDVERTLLFVLDWMOONSIFSAQ 192
QY 178 DEAKFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAIT 237
Db 178 DEAKFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAIT 237
QY 193 HQARKFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAIT 252
Db 193 HQARKFTFNMAKHMSMOPGEETEQLKKEVYTFPMKGVSAPLNPGTAYHKALOSRAIT 252
QY 238 ILKFTERMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNL 297
Db 238 ILKFTERMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNL 297
QY 253 ILGVIERMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNL 291
Db 253 ILGVIERMEERKLDIKEEDQEEVEVTEDEAKMSKSDHYRKORTDDDLGWLKHSNL 291
QY 298 TEQILDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELN 357
Db 298 TEQILDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELN 357
QY 292 KEQILDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELN 351
Db 292 KEQILDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHELEIARAKKEGESELN 351
QY 358 MDDVKKMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDS 417
Db 358 MDDVKKMDFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDS 417
QY 352 WEDYKEVFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDS 411
Db 352 WEDYKEVFTQCVINETLRGNVYRFLHRRALDVRKGYDIPSGMKVLPVISAHLNDS 411
QY 418 RYDQPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIH 477
Db 418 RYDQPNLFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIH 477
QY 412 LYEDPQNFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIH 464
Db 412 LYEDPQNFNPRWQOQNGSSGSGSFSTWGNKYNMFGGPRLCAGSELAKLEMAVFIH 464
QY 478 HLVLKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 512
Db 478 HLVLKFNWELAEDDQPFAPFPVDPNGLPIRVSRIL 512
QY 465 HLVLNFRWELAEDDQPFAPFPVDPNGLPIRVSRIL 499
Db 465 HLVLNFRWELAEDDQPFAPFPVDPNGLPIRVSRIL 499

RESULT 4

US-10-424-599-282744
; Sequence 282744, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424, 599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 282744
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_97339C.1.pep
US-10-424-599-282744

Query Match 43.5%; Score 1166.5; DB 6; Length 300;
Best Local Similarity 73.6%; Pred. No. 8.1e-105;
Matches 234; Conservative 26; Mismatches 35; Indels 23; Gaps 2;

OY 8 TLPLLLPLSLSL--LELLKRRNRKTRFNLPCKSGMPLGETIGYLPATTLG 65
DB 4 SLTFYSLSAILALLPITIFILIRKOSKPRILNPPGMMGMPFJGTLKPYSAITVG 63
OY 66 DFMOOHVSRYKIRYNSNLFGEPTVSADAGLNRFILONEGRLEFECSPRSIGILGKMS 125
DB 64 EFMEOHIRAGTYIKSKLFAGPAIVSADAGLNRFILONEGRLEFECSPRSIGILGKMS 123
OY 126 LVLVGDMDRMRISLNLNLSHARLRTILKDYERTFLVYDSWQONSIFSQDEKKTTF 185
DB 124 LVLVGDMDRMRISLNLNLSHARLRTILKDYERTFLVYDSWQONSIFSQDEKKTTF 183
OY 186 NLMAKHMSMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFIERK 245
DB 184 NLMAKHMSMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFIERK 243
OY 246 MEERKLDIKEEDQEEVEVTEDEAEMSKSDHYRKQRTDDLLGWLKHSNSTEQIIDL 305
DB 244 MEERKLDIKEEDQEEVEVTEDEAEMSKSDHYRKQRTDDLLGWLKHSNSTEQIIDL 282
OY 306 LSLFAGHETSSVAIALA 323
DB 283 LSLFAGHETSSVAIALA 300

RESULT 5
US-10-425-114-72982
; Sequence 72982, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72982
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROB73080G04_FLI.pep
US-10-425-114-72982

Query Match 43.4%; Score 1164; DB 6; Length 317;
Best Local Similarity 67.1%; Pred. No. 1.5e-104;
Matches 222; Conservative 36; Mismatches 41; Indels 32; Gaps 2;

OY 182 KFTFNLMKHMISMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFI 241
DB 3 QFTFNLMKHMISMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFI 62
OY 242 IERMEERKLDIKEEDQEEVEVTEDEAEMSKSDHYRKQRTDDLLGWLKHSNSTEQI 301
DB 63 IERMEERKLDIKEEDQEEVEVTEDEAEMSKSDHYRKQRTDDLLGWLKHSNSTEQI 101
OY 302 LDLLSLFAGHETSSVAIALAIFLQACPKAVELEPREHLEIRAKKEGESELMNDY 361
DB 102 LDLLSLFAGHETSSVAIALAIFLQACPKAVELEPREHLEIRAKKEGESELMNDY 161
OY 362 KMDFTQCVINETRLGNVRFRLHRAKALDVRVGYDIPSGMYLPVISAHLNDSRYDQ 421
DB 162 KMDFTQCVINETRLGNVRFRLHRAKALDVRVGYDIPSGMYLPVISAHLNDSRYDQ 221
OY 422 PNLFPWFMQOQNGASSGSGFSSTWGNNTMPFGGPRCAGSELAKLEMAVFIHLVL 481
DB 222 PSRFNPMWKLQSNHAPS-----FMPYGGGPRCAGSELAKLEMAVFIHLVL 270
OY 482 KFMNELAEDDQFPAPFPVDFPNGLPYRSRI 512
DB 271 KFMNELAEDDQFPAPFPVDFPNGLPYRSRI 301

RESULT 6
US-10-424-599-178693
; Sequence 178693, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424, 599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 178693
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_132377C.1.pep
US-10-424-599-178693

Query Match 37.3%; Score 1001; DB 6; Length 488;
Best Local Similarity 42.1%; Pred. No. 2.1e-88;
Matches 210; Conservative 90; Mismatches 159; Indels 40; Gaps 9;

OY 16 PSLSLFLFL--KRNKRTRFNLPPCKSGMPLGETIGYLPATTLGDPNQOYVS 73
DB 17 PSLSLFLFL--KRNKRTRFNLPPCKSGMPLGETIGYLPATTLGDPNQOYVS 76
OY 74 KYGKIRSNLFGEPITVSADAGLNRFILONEGRLEFECSPRSIGILGKMSVLVGDH 133
DB 77 KYGKIRSNLFGEPITVSADAGLNRFILONEGRLEFECSPRSIGILGKMSVLVGDH 136
OY 134 RDMRSISLNLNLSHARLRTILKDYERTFLVYDSWQONSIFSQDEKKTTFNLMKHM 193
DB 137 RDMRSISLNLNLSHARLRTILKDYERTFLVYDSWQONSIFSQDEKKTTFNLMKHM 194
OY 194 SMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFIERKMEERKLDI 253
DB 195 SMDPGEETEQLKKEYTFPMKGVASAPLNPATYHKAQSRATILKFIERKMEERKLDI 250
OY 254 KEEDQEEVEVTEDEAEMSKSDHYRKQRTDDLLGWLKHS--NLSTEQIIDLILSLFAG 312

```

Db 251 KESTTEKK-----NDMLGALLASGYHFSDEIVPMALLVAG 289
QY 313 HETSSVAIALAIFLQACPKAVELEAREHELEIAARAKKEGSELNMDYKMDTOCVIN 372
Db 290 YETSTIMTLAIKFLTEPLALQAKHEHDI-RAKSCPEAPLEMDYKMATQOCVYN 348
QY 373 ETLRLGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 432
Db 349 ETLRVANIIAIFRRAMTDINIKGYTIPKGRVVASFPAVHLNPDHEDARTFNPWRV-Q 407
QY 433 QONNGASSGSGSFSTGNNVMPGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAEDDO 492
Db 408 SNSASSP-----GNVTPFGGPRLCAGYELARVALSVFLHRTVTRISWPEAEDEK 459
QY 493 PFAFPVDFPNGLPIRVSR 511
Db 460 LVFEPTTRTQKRYPIYKR 478

```

```

RESULT 7
US-09-502-426a-19
; Sequence 19, Application US/09502426A
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sungbwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426A
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-502-426a-19

```

```

Query Match 36.5%; Score 978.5; DB 5; Length 472;
Best Local Similarity 41.9%; Pred. No. 3.2e-86;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLFLFLILKRRNRKTRFNLPRGSGMPFLGETIGYLRKYATATTGDFMOQH 71
Db 7 LLLLSIAAGFL--LLRRTYRRMGLPFGSLGLPLIGETPQLIGAYKTENPEPFIDER 63
QY 72 VSKYGIKYSNLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGGILGKMSMLVYGD 131
Db 64 VARGSVFMTHLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGGILGKMSMLVYGD 123
QY 132 MHRDMSISLNFSLHARLITLKDVRHLLFVLDMSQONSIFSADDEAKKFTFNMAKH 191
Db 124 LHRMHSLLTMSFANSSTIKDHLMDIDLVRFNLDMSQONSIFSADDEAKKFTFELTVKQ 181
QY 192 IMSNDPGEETEQOLKREYVTEFMKGVASAPLNLPGTAHKALOSRAITILKFERKMEERKL 251
Db 182 LMSFDPG-EMSESLRKEYLLVIEGFSPLPLFSTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDQEEBEVTEDEAEMSKSDHVRKQRPDDDLGLGVNLKHSNLSIQDILDLISLFA 311
Db 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIYDFLVALLYA 278
QY 312 GHETSSVAIALAIFLQACPKAVELEAREHELEIAARAKKEGSELNMDYKMDTOCVI 371
Db 279 YETSTIMTLAVFLETPLALQAKHEHEKIRAMKSD--SYSEMSDYKSMPTQCVV 336
QY 372 NETLRIGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 431

```

```

Db 337 NETLRVANIIIGVFRAMTDVIEIKYKIPKCKVYSSFRVHLDPNHEDARTFNPWRQ 396
QY 432 QONNGASSGSGSFSTGNNVMPGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAE 490
Db 397 -----SNSVTTGPSNVFTPPGGPRLCAGYELARVALSVFLHRLVTFGSVPAQ 446
QY 491 DQFAFPVDFPNGLPIRVSR 511
Db 447 DKLVFPTTRTQKRYPIYKR 467

```

```

RESULT 8
US-09-502-426b-19
; Sequence 19, Application US/09502426B
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sungbwa
; APPLICANT: FELDMANN, Kenneth A.
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 11696-070001
; CURRENT APPLICATION NUMBER: US/09/502,426B
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-502-426b-19

```

```

Query Match 36.5%; Score 978.5; DB 5; Length 472;
Best Local Similarity 41.9%; Pred. No. 3.2e-86;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLFLFLILKRRNRKTRFNLPRGSGMPFLGETIGYLRKYATATTGDFMOQH 71
Db 7 LLLLSIAAGFL--LLRRTYRRMGLPFGSLGLPLIGETPQLIGAYKTENPEPFIDER 63
QY 72 VSKYGIKYSNLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGGILGKMSMLVYGD 131
Db 64 VARGSVFMTHLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGGILGKMSMLVYGD 123
QY 132 MHRDMSISLNFSLHARLITLKDVRHLLFVLDMSQONSIFSADDEAKKFTFNMAKH 191
Db 124 LHRMHSLLTMSFANSSTIKDHLMDIDLVRFNLDMSQONSIFSADDEAKKFTFELTVKQ 181
QY 192 IMSNDPGEETEQOLKREYVTEFMKGVASAPLNLPGTAHKALOSRAITILKFERKMEERKL 251
Db 182 LMSFDPG-EMSESLRKEYLLVIEGFSPLPLFSTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDQEEBEVTEDEAEMSKSDHVRKQRPDDDLGLGVNLKHSNLSIQDILDLISLFA 311
Db 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIYDFLVALLYA 278
QY 312 GHETSSVAIALAIFLQACPKAVELEAREHELEIAARAKKEGSELNMDYKMDTOCVI 371
Db 279 YETSTIMTLAVFLETPLALQAKHEHEKIRAMKSD--SYSEMSDYKSMPTQCVV 336
QY 372 NETLRIGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 431
Db 337 NETLRVANIIIGVFRAMTDVIEIKYKIPKCKVYSSFRVHLDPNHEDARTFNPWRQ 396
QY 432 QONNGASSGSGSFSTGNNVMPGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAE 490
Db 397 -----SNSVTTGPSNVFTPPGGPRLCAGYELARVALSVFLHRLVTFGSVPAQ 446
QY 491 DQFAFPVDFPNGLPIRVSR 511
Db 447 DKLVFPTTRTQKRYPIYKR 467

```

RESULT 9

US-10-395-463-24

Sequence 24, Application US/10395463.

GENERAL INFORMATION:

APPLICANT: Tanaka, Hiroshi

APPLICANT: Kayano, Toshiaki

APPLICANT: Matsuo, Makoto

APPLICANT: Sakamoto, Tomoaki

APPLICANT: Iwahori, Shuichi

TITLE OF INVENTION: Method of Controlling Character of Monocytogenase by Modification

TITLE OF INVENTION: Overexpression of Cytochrome P450 Monooxygenase Gene Involved in

FILE REFERENCE: 59150-8022

CURRENT APPLICATION NUMBER: US/10/395,463

PRIOR FILING DATE: 2003-03-20

PRIOR APPLICATION NUMBER: JP 2002-276398

NUMBER OF SEQ ID NOS: 38

SOFTWARE: PatentIn version 3.1

SEQ ID NO 24

LENGTH: 501

TYPE: PR

ORGANISM: Oryza sativa

US-10-395-463-24

Query Match 31.9%; Score 856.5; DB 6; Length 501;

Best Local Similarity 36.9%; Pred. No. 2.6e-74;

Matches 191; Conservative 97; Mismatches 189; Indels 41; Gaps 11;

QY 11 PULLPSLILFLILK-----RRNRKTRFNLPGKSGMPLIGETIGYLRKY 59
 DB 5 PVLIAAAAVVAVVAVLRLWLLILGPAAGRLGK--RAMPGSTPLRIGETLRISXY 62
 QY 60 TATLGDPMOHSKSKYRSLRPEPTIVSADAGLNRFILQNGRLFCSTPRSIGTI 119
 DB 63 KIPNPEPTIDENYARAGVFTTHVEGETVFSADPFRNLILAAEGRAVHSSYSSIALT 122
 QY 120 LGRMSLVLVGDMHRMRSISLFLSHARLRTLLKDVRRHTLVYDLSWQNSIPSAOE 179
 DB 123 LGRSLLTRGAHKLHSLTLRLGR-ASRPLAHIDLVLATYRQNEPAPVYRLME 181
 QY 180 AKKFTNLMAKHIMSMDPEEETEDLKEEYTFMGVSAPLN---LPGTAYHKALOS 235
 DB 182 AKKFTNLMVQVYIEBQ-PWTESIIRREYKILIDGFISIPPLANILPFTYGGALKAR 240
 QY 236 ATILFIERKMEERKLDIKEEDQEEVEKTEDEAKMSDHYRKQRTDDLLGWLKHSN 295
 DB 241 KKVAGALREVYKR-----MEERKNGSGIGDEGKREKKNV-----EELL--EAEQGS 288
 QY 296 LSTEQILDILISLIFAGHETSSVAIALAIFLOACPKAVEELREHLEFARAKKEGSE 355
 DB 289 FSEEMVDFCISLIVAGYETTSMLMTLAVKFLTEPRLALAEKEHANTIRMKK--KOP 346
 QY 356 LMDYKMDFTQCVINETLRLGNVYRFLHRAKALDVRKGYDIPSGMKVLPVISAVHLD 415
 DB 347 LEMSYKSMPTQCVINETLRLGNVYRFLHRAKALDVRKGYDIPSGMKVLPVISAVHLD 406
 QY 416 NSRYDQPLNFWRMQOONNGASSSGSFTWGNVWPFGGGPRLCAGSELAKEMAVF 475
 DB 407 NEHYENARTFNPWRQINNKLNANVA-----NIFPFGGPRLCAGSELARVVSIF 459
 QY 476 IHHVLKTNMELAEEDQPFAPFVDFPGLPIRVSRI 513
 DB 460 LHHVTRFSWETEDRLVFFPTTTLKGYPIINL-RL 496

RESULT 10

US-09-502-426A-25

Sequence 25, Application US/09502426A

GENERAL INFORMATION:

APPLICANT: Azpiroz, Ricardo

APPLICANT: CHOE, Sunghwa
 APPLICANT: FELDMANN, Kenneth
 TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
 FILE REFERENCE: 2225-0001
 CURRENT APPLICATION NUMBER: US/09/502,426A
 PRIOR FILING DATE: 2000-02-11
 PRIOR APPLICATION NUMBER: 60/119,657
 PRIOR FILING DATE: 1999-02-11
 PRIOR APPLICATION NUMBER: 60/119,658
 PRIOR FILING DATE: 1999-02-11
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 25
 LENGTH: 575
 TYPE: PR
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Consensus sequence
 NAME/KEY: VARIANT
 LOCATION: (1)...(575)
 OTHER INFORMATION: Xaa = Any Amino Acid or No Amino Acid
 US-09-502-426A-25

Query Match 31.7%; Score 849; DB 5; Length 575;

Best Local Similarity 40.6%; Pred. No. 1.7e-73;

Matches 215; Conservative 35; Mismatches 242; Indels 38; Gaps 10;

QY 15 LPSILSLFLILKRR-----NKRTRFNLPGKSGMPLIGETIGYLRKYATTL 64
 DB 39 LLSXALXVXKLARARXXXXXXRRKRLPGMTGLRIGETIQFLKXXXXXX 98
 QY 65 GPMOHSKYG---KIRSNLFGPTIVSADAGLNRFILQNGRLFCSTPRSIGTI 120
 DB 99 GPMOHSKYG---KIRSNLFGPTIVSADAGLNRFILQNGRLFCSTPRSIGTI 157
 QY 121 GKMSTLVYVGDHMRMRSISLFLSHARLRTLLKDVRRHTLVYDLSW---QONSITSA 176
 DB 158 GKMSTLVYVGDHMRMRSISLFLSHARLRTLLKDVRRHTLVYDLSW---QONSITSA 217
 QY 177 QDEAKFTNLMAKHIMSMDPEEET--EOLKEEYTFMGVSAPLNPGTAYHKALOS 234
 DB 218 XXEXKMPEDXXXXXXMGXPRXXETXXXXXLYXEXKILKGLFSLPILNPTAYKALXA 277
 QY 235 RATIIFIERKMEERKLDIKEEDQEEVEKTEDEAKMSDHYRKQRTDDLLGWL--- 291
 DB 278 RAFTXAXLXXXXXXXXXXIXEXRREEXXXXXXXSXXXXXXXXXXXXDGLLXXX 337
 QY 292 -----KHSNLTSTQILDILISLIFAGHETSSVAIALAIFLOACPKAVEELREHLEI 344
 DB 338 XXXXDEXXXLSXDEIXDXLXLLIFAGHETSSXLAANVFLXHPVXEXLREHXA 397
 QY 345 ARAKKEGESELMNDYKMDFTQCVINETLRLGNVYRFLHRAKALDVRKGYDIPSGMK 404
 DB 398 XRAKXXXESLTXDXDKKXKXVIVINETLRLXIVGKPRXAKKDVINGIYAIKPKWK 457
 QY 405 VLPVISAHLNDSKRDQNLNFWRMQOONNGASSSGSFTWGNVWPFGGGPRLCAG 464
 DB 458 VYSTRVAVHLDPDYKDEKFNPKWXXXXXXSXXXXXXXXXXXXXNXXPFGGGPRLCAG 517
 QY 465 SELAKEMAVFIIHVLKTNMELAEEDQPFAPFVDFP-----NGLPIRV 509
 DB 518 KEMAKLEKXVFLHRLVQXF-WELAXXND-XXXXLVXFTTXXRPXNDLPIKV 565

RESULT 11

US-09-502-426B-25

Sequence 25, Application US/09502426B

GENERAL INFORMATION:

APPLICANT: Azpiroz, Ricardo

APPLICANT: Choe, Sunghwa

APPLICANT: Feldmann, Kenneth A.

TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF

FILE REFERENCE: 11696-070001
CURRENT APPLICATION NUMBER: US/09/502.426B
CURRENT FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/119,657
PRIOR FILING DATE: 1999-02-11
PRIOR APPLICATION NUMBER: US 60/119,658
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 25
LENGTH: 575
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus sequence
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(575)
OTHER INFORMATION: Xaa = Any Amino Acid or No Amino Acid
US-09-502-426B-25

Query Match 31.7%; Score 849; DB 5; Length 575;
Best Local Similarity 40.6%; Pred. No. 1.7e-73;
Matches 215; Conservative 35; Mismatches 242; Indels 38; Gaps 10:
QY 15 LPSLISLLFLILIKR-----NRKTRFNLPGKSGWPLGETIGYLRKYATTL 64
DB 39 LLSXALXVXLXLAARXXXXXXKXXXXXXRRKXLPQTMGLPYLGETLOFLKXXXXXP 98
QY 65 GDFMOQHVSKYK---KTYRSNLGEPITVSADAGLNFLQNGS-RLFECYPRSIGIL 120
DB 99 GDFPKKENVXXGXXXXXXIK-HLSEPTIXSDAELNPKLXNEGKLEFCXKXPSXKXL 157
QY 121 GKSMVLVGDMDHMRMSISLNFLSHARLRTILKDYERHTLFVLDG---QONSIFSA 176
DB 158 GKSLXAXXGKXEHKRMXLLXSXFSXXXLDHXLPIIDRXVRSKXKXKXKXKXKXIXX 217
QY 177 QDEKKTFTNLMKHMISMDGEEET--EQLKREYVFMKGVSAPLNPGTAYHKALQS 234
DB 218 XHEKKTFTFXKX 277
QY 235 RATLKFIERMEERKLDIKEEDQEEEVTEDEAEMSKSDHVKORITDDLLGMVL--- 291
DB 278 RAFAKXALXXXXXXXXXXXXXIXEKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 337
QY 292 -----KHSNLSQEIIDLILSLFAGHETSVALALAFLOACPRVAEELREHLEI 344
DB 338 XXXXEDXXXXLSDXEXIXDXIXLLFAGHETSSXKXAVFLKHPDVXKXLEHXA 397
QY 345 ARAKKEGESELMNDYDKKDFTOCVINELRLGNVYRFLHKKALKDYRKGYDIPSGWK 404
DB 398 XRAKX 457
QY 405 VLPIYAVHLDNSRYDQNLFPWRMOQNNGASSGSGSFSTGNVMPMGSGPRICAG 464
DB 458 VXSIRAVHLDPOXYPRPEKFNPRKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 517
QY 465 SELAKLEMAVFIIHLLVLFKFWELAEDDQPAFPFVDEP-----NGLPYRV 509
DB 518 KEIAKLEMAVFILRLVQXF-WELAXXXD-XXXXLVXEPTRKXNLLIKV 565

RESULT 12
US-10-310-154-384
Sequence 384, Application US/10310154
GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
APPLICANT: Chomec, Paul S.
APPLICANT: Adams, Thomas H
APPLICANT: Ruff, Thomas G.
APPLICANT: Agarwal, Ameeta K.
APPLICANT: Ahrens, Jeffrey E.
APPLICANT: Ball, James A.

APPLICANT: Banu, G.
APPLICANT: Bell, Erin
APPLICANT: Boddapalli, Raghava
APPLICANT: Delkman, Jill
APPLICANT: Deng, Molian
APPLICANT: Dong, Jintuo
APPLICANT: Duff, Stephen M.
APPLICANT: Galligan, Meghan M.
APPLICANT: Hinchey, Brenda S.
APPLICANT: Huang, Shishleh
APPLICANT: Johnson, G. Richard
APPLICANT: Jung, Vincent
APPLICANT: Kretzmer, Keith A
APPLICANT: Laccetti, Lucille B.
APPLICANT: Lai, Chao-Qiang
APPLICANT: Lee, Gary
APPLICANT: Liu, Jie-Yi
APPLICANT: Liu, Jindong
APPLICANT: Lu, Bin
APPLICANT: Luethy, Michael M.
APPLICANT: Lund, Adrian
APPLICANT: Madson, Linda L.
APPLICANT: Malloy, Kathleen A.
APPLICANT: McKiel, Christine L.
APPLICANT: Miller, Philip W.
APPLICANT: Padmavathi, Manchi Kant
APPLICANT: Parnell, Laurence D.
APPLICANT: Sten, William G.
APPLICANT: Tennessee, Dan
APPLICANT: Vidya, K.R.
APPLICANT: Wang, Haiyuo
APPLICANT: Xin, Zhanqun
APPLICANT: Xu, Nanfei
APPLICANT: Yang, Chunzhi
APPLICANT: Zeng, Xiaoping
APPLICANT: Zhang, Qiang
APPLICANT: Zhao, Yajuan
APPLICANT: Zhou, Li
TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
FILE REFERENCE: 38-15(52796)B
CURRENT APPLICATION NUMBER: US/10/310,154
CURRENT FILING DATE: 2002-12-04
PRIOR APPLICATION NUMBER: 60/337,358
PRIOR FILING DATE: 2001-12-04
NUMBER OF SEQ ID NOS: 736
SEQ ID NO 384
LENGTH: 503
TYPE: PRT
ORGANISM: Zea mays
US-10-310-154-384

Query Match 31.0%; Score 832; DB 6; Length 503;
Best Local Similarity 37.2%; Pred. No. 6.2e-72;
Matches 191; Conservative 95; Mismatches 191; Indels 36; Gaps 12:
QY 11 PLLPISLISLL-FLILIKR-----NRKTRFNLPGKSGWPLGETIGYLRKYATTLG 65
DB 8 PLFLIGAVLALALGLVWVRAAGRRASARLPFGSTGPLIGETIRLLIAAYKSPNPE 67
QY 66 DFMQHVSKYK-ITYRSNLGEPITVSADAGLNFLQNGS-RLFECYPRSIGILGKWS 124
DB 68 PFIDERVARRHSGVFTTHVGEERTVFSADPAFNRLLLAEGRAVASCYSPSSIAVLLGPRS 127
QY 125 MLVYVGDMDHMRMSISLNFLSHARLRTILKDYERHTLFVLDGMOQNSIFSAQ--DEAKK 182
DB 128 LLITSGFAHKLHSLITLRLGRP-ASPLLAHIDRLVATYRGMDBPAGMARLLDEK 186
QY 183 FTENIMAKHMISMDPGEETEQLKREYVFMKGVSAPLN---LPGTAYHKALQSRTI 238
DB 187 ITFMVLVQWLVASIBG-PWTFESVREVYKLVDFGSIFPPASLLPFTVYGOALKARKKV 245
QY 239 LKFERMEERKLDIKEEDQEEEVTEDEAEMSKSDHVKORITDDLLGMVLKHSNLS 298

Db 246 AGALREYVRKMD-----KAEDDCASKKNGEKKDVEELL---EARGSPSV 291
QY 299 EQLDILLSLFAGHETSSVAIALAIFLQACPAVEELREHLEIARAKLEDESELNW 358
Db 292 EEWDFGLSLVAVETSTVMTLAVKFLTETPTALMOLEHDSI-RGYAGKQA-Lew 349
QY 359 DDYKMOFTQCVINETRLGNVRELRKALKDVRYKGYDIPSGKVLPIVSAVHLNDR 418
Db 350 SDYKMPETQCVISETLRVANLISGVFRANTDHFHGVHVPKCKTIFASFRVHLSIDH 409
QY 419 YDQPNLFNPMWQOONNGASSGSGSFSTWNNMPPGGGPRCLCAGSELAKLENAVEIHH 478
Db 410 YEAKRTDPMWQOGKSKLOSAGASLFT-----PPGGGPRCLCPGYELARVYVSLHH 463
QY 479 LVLENNELAEDDQFAFPEVDFPNGIPRYSR 511
Db 464 LVTRFSWEAEEDRIVEFTTTRILKGPILLRR 496

RESULT 13

US-10-356-153-92
Sequence 92, Application US/10356153
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
FILE REFERENCE: 62773
CURRENT APPLICATION NUMBER: US/10/356,153
CURRENT FILING DATE: 2003-01-29
PRIOR APPLICATION NUMBER: US/10/142,231
PRIOR FILING DATE: 2002-05-08
PRIOR APPLICATION NUMBER: 60/165,250
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 92
LENGTH: 486
TYPE: PRT
ORGANISM: *Taxus cuspidata*
US-10-356-153-92

Query Match 30.5%; Score 816.5; DB 6; Length 486;
Best Local Similarity 35.1%; Pred. No. 1.9e-70;
Matches 176; Conservative 97; Mismatches 184; Indels 45; Gaps 8;
QY 19 LSLILFLILKRRNK-----TRFNLPPGKSGWPLGTIGTILKPYATTGDFWQCHV 72
Db 20 LFTCVTLILRRSNDROGNSANKPLPPGSAGLPPIGTITFLINDAKSPGRKKPDEHE 79
QY 73 SKYKTIYRSNLFGEPT-TIVSADAGINRFTLQNEGRLEFCSYRSIGGILGKMSMLVQGD 131
Db 80 LKYGPIFGSLGRTRAVVSDPERKYYLQNEGRLEFESNALAPPNNLIGKYGSLAVQGE 139
QY 133 MHRDMSISLNFLSHARLTIILKDERHTLFLVDSWOONSJFSADDEAKKRTFPMIMKH 191
Db 140 LQRLHATVAVNLKHEHTLSSDFMEDIDIFQAGMKWKEEGIPLOHKNOVLVNDMAR 199
QY 192 IMSMPGEETOLKKEVYTPKGYVAPLNPGLVYHKAQSRATTIILKTERKKEERLT 251
Db 200 LLDLPP-SEMGHITKAFDDFVGAVALSEPLNPPTVARGIARGLILKRIHKICKERR- 257
QY 252 DIKEDEEEVEKTEDEAEMSKSDHVRKORTDDLLGWLKHNSNSTQOIDLILSLFA 311
Db 258 -----BHEPV-----LRNDLLTLVREGTFSDELITDITIFEFA 292
QY 312 GHETSSVALAIFLQACPAVEELREHLEIARAKLESEELNMDYKMDFTQCVI 371
Db 293 GVEISAMAMTEAVKTLAENPRALBELRAEHDLALKAKG-GNEKLTWMDYQSMKRVHCVI 351
QY 372 NETTLGNVRELRKALKDVRYKGYDIPSGKVLPIVSAVHLNDRIDQPNLEPNWQ 431
Db 352 NETLRLGATVYLFREAKODITKADDEVLPKGTWVSFVLSAHHYDGKHYEADKFLPMWQ 411
QY 432 QONNGASSGSGSFSTWNNMPPGGGPRCLCAGSELAKLENAVEIHHVLKFNELAEDD 491

Db 412 NE-----GOETLEPCYMPGRCGRILCPGILHARFIALFHNFTKFRWOLEID 462
QY 492 QPFAFPVDFPNGIPRYSR-SRI 512
Db 463 RATYFPLPSTENGPIRILYSRV 484

RESULT 14

US-10-310-154-383
Sequence 383, Application US/10310154
GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
APPLICANT: Chomet, Paul S.
APPLICANT: Adams, Thomas H
APPLICANT: Ruff, Thomas G.
APPLICANT: Agarwal, Ameeta K.
APPLICANT: Ahrens, Jeffrey E.
APPLICANT: Ball, James A.
APPLICANT: Bannu, G.
APPLICANT: Bell, Erin
APPLICANT: Boddupalli, Raghava
APPLICANT: Deikman, Jill
APPLICANT: Dong, Molian
APPLICANT: Duff, Stephen M.
APPLICANT: Galligan, Meghan M.
APPLICANT: Hinchey, Brenda S.
APPLICANT: Huang, Shihshieh
APPLICANT: Johnson, G. Richard
APPLICANT: Jung, Vincent
APPLICANT: Kretzmer, Keith A
APPLICANT: Laccetti, Lucille B.
APPLICANT: Lai, Chao-Qiang
APPLICANT: Lee, Gary
APPLICANT: Lin, Jie-Yi
APPLICANT: Liu, Jindong
APPLICANT: Lu, Bin
APPLICANT: Luethy, Michael M.
APPLICANT: Lund, Adrian
APPLICANT: Madison, Linda L.
APPLICANT: Malloy, Kathleen A.
APPLICANT: McKiel, Christine L.
APPLICANT: Miller, Philip W.
APPLICANT: Padnavathi, Manohkanti
APPLICANT: Parnell, Laurence D.
APPLICANT: Start, William G.
APPLICANT: Tennesen, Dan
APPLICANT: Vidya, K.R.
APPLICANT: Wang, Haiyun
APPLICANT: Wang, Zhanguo
APPLICANT: Xu, Nanrel
APPLICANT: Yang, Chunzhi
APPLICANT: Zeng, Xiaoping
APPLICANT: Zhang, Qiang
APPLICANT: Zhao, Yajuan
APPLICANT: Zhou, Li
TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
CURRENT APPLICATION NUMBER: US/10/310,154
PRIOR FILING DATE: 2002-12-04
PRIOR APPLICATION NUMBER: 60/337,358
PRIOR FILING DATE: 2001-12-04
NUMBER OF SEQ ID NOS: 736
SEQ ID NO 383
LENGTH: 518
TYPE: PRT
ORGANISM: *Zea mays*
US-10-310-154-383

Query Match 29.5%; Score 792; DB 6; Length 518;
Best Local Similarity 33.0%; Pred. No. 5.1e-68;
Matches 160; Conservative 121; Mismatches 162; Indels 42; Gaps 9;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:55:23 ; Search time 29 Seconds

(without alignments)
1700.585 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681

Sequence: 1 MFETEHHTLPLLLPLLS.....FAFPVDEPNGLPIRVSRIL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2677	99.9	513	2 T46143	steroid 22-alpha-h
2	978.5	36.5	472	1 S55379	cytochrome P450 cy
3	866	33.3	512	2 H96759	probable steroid 2
4	860	32.1	457	2 D85429	cytochrome P450 11
5	815	30.4	382	2 T48613	hypothetical prote
6	761.5	28.4	464	2 T07859	cytochrome P450 ho
7	632	23.6	457	2 T04444	cytochrome P450 -
8	611	22.8	482	1 T02739	probable cytochrom
9	581	21.7	487	1 C71417	cytochrome P450 dl
10	579.5	21.6	485	2 A84859	probable cytochrom
11	579.5	21.6	490	2 H86185	hypothetical prote
12	577	21.5	455	2 T48973	cytochrome P450-11
13	547.5	20.4	489	2 B84733	probable cytochrom
14	497.5	18.6	460	2 D96813	hypothetical prote
15	491.5	18.3	519	1 T02263	cytochrome P450 DW
16	489.5	18.3	255	2 T04602	cytochrome P450 ho
17	479.5	17.9	444	1 S75761	cytochrome P450 -
18	369	13.8	349	2 A86329	FLAP1.4 protein -
19	357	13.3	518	2 T20908	hypothetical prote
20	338	12.6	453	2 C83722	cytochrome P450 hy
21	329.5	12.3	517	2 T20907	hypothetical prote
22	324.5	12.1	520	2 T24778	hypothetical prote
23	321.5	12.0	518	2 T24783	hypothetical prote
24	316	11.8	500	2 T04737	cytochrome P450 ho
25	310.5	11.6	520	2 T24777	hypothetical prote
26	310	11.6	500	2 T25175	cytochrome P450 mo
27	306.5	11.4	504	2 A25222	cytochrome P450 3A
28	302.5	11.3	491	2 S31277	cytochrome P450 2B
29	300	11.2	491	2 T84735	testosterone 16a-h

30	299.5	11.2	492	2 S27160	cytochrome P450 2B
31	297	11.1	491	2 A31047	testosterone 16a1p
32	297	11.1	494	2 A33293	cytochrome P450 2A
33	294.5	11.0	491	1 O4RBP	cytochrome P450 2B
34	294	11.0	506	2 D96872	probable Cytochrom
35	294	11.0	516	2 T48140	flavonoid 3',5'-hy
36	292	10.9	491	1 O4RBP	cytochrome P450 2B
37	291.5	10.9	491	2 S35666	cytochrome P450 2B
38	291.5	10.9	530	1 A27491	lanosterol 14alpha
39	290.5	10.8	502	2 T05246	cytochrome P450 mo
40	288.5	10.8	504	2 A29410	cytochrome P450, g
41	288	10.7	459	2 AG1987	cytochrome P450 11
42	287.5	10.7	520	2 T24780	hypothetical prote
43	287	10.7	503	2 JC4702	cytochrome P450 3A
44	284.5	10.6	505	1 S38534	cytochrome P450 76
45	264	10.6	501	2 T04735	cytochrome P450 ho

ALIGNMENTS

RESULT 1

T46143

steroid 22-alpha-hydroxylase (DWF4) - Arabidopsis thaliana

N:Alternate names: protein T3A5.40

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001

C/Accession: T46143

R:Biocker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.; Quettier, F.

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223024

A/Accession: T46143

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1513 <BLO>

A/Cross-references: EMBL:AL132979

A/Experimental source: cultivar Columbia; BAC clone T3A5

C:Genetics:

A:Map position: 3

A:Insertions: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3

A>Note: T3A5.40

C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology

C/Keywords: heme; iron; metalloprotein

F:308-484/Domain: cytochrome P450 homology <P45>

F:462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 99.98; Score 2677; DB 2; Length 513;

Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MFETEHHTLPLLLPLLSILFLILKRNKTRPNLPGRSGMFLGTTGYLKPPT	60
DB	1	MFETEHHTLPLLLPLLSILFLILKRNKTRPNLPGRSGMFLGTTGYLKPPT	60
QY	61	ATTLGDMQOHVSKYKIRYSNLFGEPTIYSADAGLNRTLLONEGRLEFCSYPSIGITL	120
DB	61	ATTLGDMQOHVSKYKIRYSNLFGEPTIYSADAGLNRTLLONEGRLEFCSYPSIGITL	120
QY	121	GKSMALVGDMDRMRSLNLSARLRTLLKDYERTLLVLSWQONSIFSADDEA	180
DB	121	GKSMALVGDMDRMRSLNLSARLRTLLKDYERTLLVLSWQONSIFSADDEA	180
QY	181	KKTFPNMAKHINSMQGEETOLKKEYTTPKGVASAPLNPGTAYHKAALOSRATILK	240
DB	181	KKTFPNMAKHINSMQGEETOLKKEYTTPKGVASAPLNPGTAYHKAALOSRATILK	240
QY	241	FIERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLGWALKSNLSTEQ	300
DB	241	FIERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLGWALKSNLSTEQ	300
QY	301	IIDLLIISLFAGHETSSVAIALAIFPLQACPKAVEELREHLEIARAKKELGSELNWD	360
DB	301	IIDLLIISLFAGHETSSVAIALAIFPLQACPKAVEELREHLEIARAKKELGSELNWD	360

QY	361	KKKMDFTQCVINETRLIGVWVEFLRRKALKDYRYKYDILPSGKXVLPVISAVALDMSRYD	420
Db	361	YKKMDFTQCVINETRLIGVWVEFLRRKALKDYRYKYDILPSGKXVLPVISAVALDMSRYD	420
QY	421	QPNLEPNRMQOONNGASSGSGSFSTWGNNTNMPFGGPRLCAGSELAKEMAVFTIHLV	480
Db	421	QPNLEPNRMQOONNGASSGSGSFSTWGNNTNMPFGGPRLCAGSELAKEMAVFTIHLV	480
QY	481	LKFNWELAEDDQPEFAFPVDFPENGPIRYSRIIL	513
Db	481	LKFNWELAEDDKPEFAFPVDFPENGPIRYSRIIL	513

RESULT 2

cytochrome P450CYP90 - Arabidopsis thaliana
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S55379
R:Seekeres, M.; Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.
submitted to the EMBL Data Library, May 1995
A:Reference number: S55379
A:Accession: S55379
A:Molecule type: mRNA
A:Residues: 1-472 <SIZE>
A:Cross-references: EMBL:X87367; NID:g853718; PIDN:CAA60793.1; PID:g853719
C:Genetics:
A:Gene: CYP90
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase
F:275-440/domain: cytochrome P450 homology <445>
F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

	36.5%	Score 978.5;	DB 1,	Length 472;
Query Match				
Best Local Similarity	41.9%	Pred. No. 7.5e-56;		
Matches 210; Conservative	79;	Mismatches 171;		
				Indels 41; Gaps 8;

```

0Y 12 ILLPSLLSTLLFLLLRNRKKTFFNNPGKSGMFFGETTGYLKPRYATTLGDMQOH 71
Db 7 LLLLSSTIAGL---LLLRTRYRMGCLPGSLGLRIGETFOOLGAYVTENDEPPTDR 63
QY 72 VSKYGIKYSNLFGEPITVSADAGLNRFLLQNEGRLLFECSPRSIGIGLKGWMLVGD 131
Db 64 VARGSVYMTLLFGEPITFSADPEPNRYVLQNEGRKFECSYASIONLLGKSHLLMKGS 123
QY 132 MHRMRSISLNFLSHARLTLLKVERHTLVLDWMOONSIFSADEAKKFTENIMAKH 191
Db 124 LHKRMHSLTMSFANSSIIKDHMLMDIDRLVRFNLDMSWSRVLV--MEEAKKITFELTVKO 181
QY 192 IMSMDPGEELTEOKKEVYTFKKGVSAPLNPGYAHKALSRATILFIEKKMEERKL 251
Db 182 IMSDPG-EMSESKLEKTLVIEGFSLPLPSTTYRAIGAR-----RKVEALT 232
QY 252 DIKEDEDEEEVKTDEDEAKMSDHYRKQRTDDLLGWLKHSNLTSEOILDLISLFA 311
Db 233 VVVMKREEEGAE-----RKMDMLAALLAADGDFDEELVDEVLAVYA 278
QY 312 GHETSSVAIALAIFFLQACPKVAVELREBHLEIARAKKELGSELNWDDYKKMDPTQCVI 371
Db 279 GYETTSTMTLAVKFLTFETPLALQKKEHETIRAMSD--SYSLEMSDYKSMFPQCVY 336
QY 372 NETLRIGVNVAFLLHKALKDVRKYGDYPSGKKYLVPIVASIVLNDNSRYOQNLFNWRMQ 431
Db 337 NETLRVAVITJGVFFRAATDVEIKYKLPKGMKVVSSFAVHLDHPHEDARTFNWRMQ 396
QY 432 QONNGASSGSGSEFTGWN--YMPGGGPRCLCAGSELAKLEAVFTHNLVLFKNWEIAED 490
Db 397 -----SNSYTTGPSNVFTPPGGGPRCLCPGIELARVALSVFLHRLVTGGSWPALD 446
QY 491 DQPAFPVDFVDPNGLPIYVS 511
I: : I I

```

Db 447 DKLVFFPTTRTQKRYPFVKR 467

RESULT 3

Probable steroid 22-alpha-hydroxylase [P9124.44 [imported]] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Apr-2002
 C:Accession: H96759
 R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; An-
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Deng,
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marila
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tejo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016712; PMID:11130712
 A:Accession: H96759
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 <STO>
 A:Cross-references: GB:AE005173; NID:g11120803; PIDN:AAG30983.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T9L24.44
 A:Map position: 1
 C:Subfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 C:455/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	32.3%	Score 866;	DB 2;	length 512;
Best Local Similarity	33.8%	Pred. No. 1.6e-48;		
Matches 181;	Conservative 105;	Mismatches 173;	Indels 76;	Gaps 7;

```

QY      12 LLILPSSLILLLFILI-----LKRNRKTFPNLIPGSGMPRLGETIGLYKPTATILGD 66
      11 LLVSSSTYFLAFITIFLLAGIARRKRRADPHRLPGSRGMPLLGDYAMANAAGSPSS 70
QY      67 FMOOHVSK-----YGIYRSNLGEPITIVSADAGLNFIL 101
      71 FVEKQIKFPLSILCSYLLLLILKKRPDNGSGFMEIKRGRIFSCSLGKMAVVSADDPFNFM 130
DY      102 QNEGRLEFCSYPNISIGGIIGKWSMLVLVGDMMHMRISILNLSHARLRTILKDYERHT 161
      131 QNEGKLEFQSSYPSKPSFRLVYKGDGYIVHGDQORLHIASMMHNDQLKTNHLEVIPIVM 190
QY      162 LFVLDMSQONSISADDEAKKTFPNLMAKIMSDPEEETEDOLKYYTFMGYVSAPL 221
      191 LQPLNSKDEEYVLLDDICRKVAIHLMNOLLGVS-SESEYDEMSQLESPFVGCISLVP 249
DY      222 NLPGTVAHKLOSATILKFTERRMEERKLDIKEEDQEEBEVYTEDDAEMSKSDHVAKOR 281
      250 DLGEGTYNKKMAKKEILTKIKINTETKRLONKASD-----T 286
QY      282 TDDDLGVWLKHSNLSTEQILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREEH 341
      287 AGNGVIGRLLEESLIPRESMADPILINLFGNETSTMTLFAVYFLHCRAKMTQLLEEH 346
QY      342 LETARAKKEGSELNWDYDKKMDFTQCVINETFLRLGNVVRFLHRAALKDVRKGYDIPS 401
      347 -----DRLAGMLTMDQYKTMDEFTQCVIDEFLRLGIAIMLIMREAKEBDSYQDYIIPK 399
DY      402 GKAVYLEVISAIVHLDNSRYDOPNLFNPRRW-----QQQNGGASSGSGSFTWGNMYPEG 456
      400 GCEVVEPLSAVHDESYKESLSFNPRMWDLPETQOKRMRTSP-----FYCPIFG 449
QY      457 GGFRLAGSELAKLEMAVFIHHLVLKPNMELASDDDPAPFVDPFGILIRVSR 511
      450 GGFRCFGALARKQIALPLHFIITYYKTKQLDEDRSTFPFSARLVNGEKTQILNR 504

```

RESULT 4

D85429

cytochrome P450 like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001

C:Accession: D85429

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: D85429

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <STO>

A:Cross-references: GB:NC_001268; NID:g7270586; PIDN:CAB80304.1; GSPDB:GM00140

A:Gene: AT4g36380

A:Map position: 4

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 32.1%; Score 860; DB 2; Length 457;

Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

39 LPPGSGMPFLGEGTGYL-----KPYATTLGDFMQOHVSKYKGRISNLGEPPIYVA 92

2 IPNGLGMPVIGETLNFACGYSSRPV-----FDKRRKSLGKFKKNIIGTPIIST 55

93 DAGLNRFLIQNEGRLEFESYSPRSIGIIGKMSMLVVGDMHDMRSISINFLSHARLRT 152

56 DAEVKKVYLQNHGNFVFPAYPKSITELGENSEILSINGPHQRIHLIGAFILRSPILKMR 115

153 LKQVREHTLFLVLDNQONSIFSADDEAKKFTFNMAKIHMSMDGEEETQKKEVYTF 212

116 ITRDIEASVYLTLSMAQOLPLVHVODEIKKMFELLVKVLMSTSPG-EDMNLILKEFEF 174

213 MKGVSAPIPLPGTAYVHKALQSRATILKFERKMEERKLDIKEEDQEEVEVKT-----E 266

175 IKGILCIPKIPGRTLYLSKAKERLIKVKVVEERQVAMTTTSANDVYVLDLGGSD 234

267 DEAEKSKDHYRKQRTDDDLGLGVLNKSNLSTEQILDLISLFGHETSVAVALAIFP 326

235 SEKQSPSDPVSFGK-----IVEMMIPGSEETMPAMTLAVKE 270

327 LQACKAYBELREHELEIARAKKEGSELNMDYKKMDFQCQVINEITRLGNVYRFLHR 386

271 LSDNFAVLAKEEENMEMKRRRLDGE-EYKWTDYMSLSTQVINEITRLMANIINGVVR 329

387 KALKDVRKGYDIPSGMKVLPYISAVHLDNSRYDOPNLFNPRMOOQNNGASSGSGSFS 446

330 KALKDVEIKGYLIPGKVCVASFIVHMDIEDIYDNPQDFPKMRIRINSANSSIC----- 385

447 TWGNNTYMPFGGCPRLCAGSELALEMAYFIHHLVLFKFWELAEEDQPFAPFPVDPNGLP 506

386 -----FTFPGGGQRLCPGLELSKLEISIFLHLVRYSW-TAEDEIVSFPPVKKMRRLP 439

507 IRVSRI 512

440 IRVATV 445

RESULT 5

T48613

hypothetical protein F18022.190 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48613

A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes

A:Reference number: 224493

A:Accession: T48613

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-382 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Insertions: 31/2; 139/3; 204/1; 224/3; 253/3; 280/1; 315/3; 349/2

A:Note: F18022.190

Query Match

Best Local Similarity 30.4%; Score 815; DB 2; Length 382;

Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

45 GWFPLEETIGYLPKYATTLGDFMQOHVSKYKGRISNLGEPPIYVADAGLNRFLIQNE 104

2 GWFPLEETISFFKPHRSDSIGTFLQORVSRKGYKFSNLCGKAVVSCDQELNMFILQNE 61

105 GRLEFESYSPRSIGIIGKMSMLVVGDMHDMRSISINFLSHARLRTILKQVREHTLFLV 164

62 GKLETSDFRAMHDLIGKTSILATGELIRKLNVIISFTNLTKSPDFLHCANLSIST 121

165 LDSMOONSIFSADDEAKKFTFNMAKIHMSMDGEEETQKKEVYTFMKGVSAPIPLP 224

122 LKSMKCRREVPEFKKFTLSVYVNOQLSINKEDPARLYVLDPLSYMKGFTSLPLP 181

225 GTRAYHALQSRATILKFERKMEERKLDIKEEDQEEVEVKTDEAEKSKDHYRKQRTDD 284

182 GGTGTVAIVRSNRNIHONALIEDMNNALREEDFLISISNED----- 224

285 DLLGWVLKNSNLSTEQILDLISLFGHETSVAVALAIFLQACKAYBELREHELEI 344

225 -----EHAAT 220

345 ARAKKEGSELNMDYKKMDFQCQVINEITRLGNVYRFLHRKALKDVRKGYDIPSGWK 404

231 -RAKKDGEL-LNWEDEYQKKEFTQCVISEALRCGNIVKTVHRAKATHDIKENEYVIPKWK 288

405 VLPYISAVHLDNSRYDOPNLFNPRMOOQNNGASSGSGSFSWGNNTYMPFGGCPRLCAG 464

289 VEPITFAVHLDPSELNPEFNPFRWTKT-----AFGGGVAVCPG 329

465 SELAKLEMAVFIIHHLVLFKFWELAEEDQPFAPFPVDPNGLP 509

330 GELGKQIAFLHLHLVLSRWKIKSDEMPIAHRYVEFKGMLEI 374

RESULT 6

T07859

cytochrome P450 homolog - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001

C:Accession: T07859

R:Bishop, G.J.; Harrison, K.; Jones, J.D.

Plant Cell 8, 959-969, 1996

A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes

A:Reference number: 216181; MUID:96266705; PMID:8672892

A:Accession: T07859

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-464 <BIS>

A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AAB17070.1; PID:g1421741

A:Experimental source: strain GCR758

C:Genetics:

A:Gene: dwarf

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

F:273-436/Domain: cytochrome P450 homology <P45>

Query Match

Best Local Similarity 28.4%; Score 761.5; DB 2; Length 464;

Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

12 LLLPLSLTLFLILKRRN--RRTRFNLPKSGSGMPFLGFTIGLKPATYATTLG-DFM 68

```

Db      5 LIFLSFFGLCIFTALLRMNOVYKQANLPDGMWGLFEGTEFEK-----LGPSEM 58
QY      69 QOHVSKYGIKTYRSMIFGEPTIVSADAGLNREFTLONEGRFECSPRSIGIGLKWSMLVL 128
Db      59 KNOBARSGSPFKSHILGPTIVSMDSELNRIILVNEAKGLVPGVPGSMIDILGKNIAAV 118
QY      129 VGDHMRMRSLNLFSLHARLRTLLKDVREHRLFLVDSMOONSIFSADAKKFTFNLM 188
Db      119 NSAHKYMKGALLSLISPTMRDLPKIDDEFMRSHLNM-DNKVIDIOEKTKMAFLSS 177
QY      189 AKHMSMDPGEET---EOLKKEYVTFMKGVSAPLNPGRAYHKAQSRATILKFERK 245
Db      178 LKQI---AGIESTSLAOEFMSEFNLVGLTSLPINLPNTNHRGFOARKIIVNLFTL 233
QY      246 MEERKLDIKEDDEEVEETDEDAEMSKSDHVRKQRTDDLLGWVLKHS---NLSTEQI 301
Db      234 IEERR-----ASKEQHDMLGYLMEARFRLTDDDM 266
QY      302 LDLLSLFAGHETSVAIALAIFLQACPAVEELREHELEIARAKKEGSELMNDY 361
Db      267 IDLTITLYSGYEVSTSMAYKYLHDHPVLEELREHMAIREKKP--EDPIDYNDY 324
QY      362 KKMDFTQCVINETRLGNVYRFLRKALKDVRKGYDIPSGKYLPIVSAVHLNDSRYDQ 421
Db      325 RSMRFTAVILERSRLATTVNGVLRKTTQDMETNGITLIPKGRIVYTRRLNDYPRD 384
QY      422 PNLFPWRMOQOONASSSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHLVL 481
Db      385 PYSFNPWRMKDS-----LEHONSFLVGGGRQCPGKEIGVAELSTFLHYVT 433
QY      482 KFMNELAEDDOPFAPFVDPNGLPIRVS 510
Db      434 KYRWEITGDKLMKFPVEAPNGLIRVS 462

```

RESULT 7

```

T04444
cytochrome P450 - Arabidopsis thaliana
N:Alternate names: protein T18B16.200; protein T5K18.10
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
C:Accession: T04444; T05806
R:Bevan, M.; Benes, V.; Reichmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A:Accession number: Z15359
A:Molecule type: DNA
A:Residues: 1-457 <BEV>
A:Cross-references: EMBL:AL021687
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15453
A:Accession: T05806
A:Molecule type: DNA
A:Residues: 131-457 <BEV>
A:Cross-references: EMBL:AL022580
A:Experimental source: cultivar Columbia; BAC clone T5K18
C:Genetics:
A:Map position: 4
A:Introns: 67/2; 173/3; 302/3; 358/1; 393/3
A:Note: T18B16.200; T5K18.10
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F:272-433/Domain: cytochrome P450 homology <P45>

```

```

Query Match      23.8%; Score 632; DB 2; Length 457;
Best Local Similarity 31.8%; Pred. No. 1.8e-33;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;
QY      13 LLLPSLLSLFLILKRRNRKTRFNLPPKSGMPLGERTIGLYKPTATTLGDFMOOHV 72

```

```

Db      10 LFAGSLFLYLKLLLSQRFPGSSKLPLPPTGMWVPEYGET---FOLYSDDP-NVFGSQ 65
QY      73 SKYGRVTSNLFGEPTIVSADAGLNREFTLONEGRFECSPRSIGIGLKWSMLVL 132
Db      66 KRYGVSFTTHVLCGCVMISPEAKFVLVTKSHLFKPFPPSKERMALGKQAIFFHQDGY 125
QY      133 HDMSRISLNLFSHARLRTLLKDVREHRLFLVDSMOONSIFSADAKKFTFNLM 192
Db      126 HAKRLKLVRAFPMSIRN-NVPDIESTIAQDSLRSW-EOTMINTVOEMKTYFNVALLSI 183
QY      193 MSMDGEET---EOLKKEYVTFMKGVSAPLNPGRAYHKAQSRATILKFERK 249
Db      184 F---GKDEVLVREDLRKCYLLLENGYSMPNLBGTTFHKMKARKELSOIARLTSR 239
QY      250 KLDIKEDQEEBEVTEDEAEMSKSDHVRKQRTDDLLG-VWLKSNLSTEQIIDLISL 308
Db      240 R-----ONGSSH-----NDLGSFNGDKBELTDEQIADNITGV 272
QY      309 LFAGHETSVAIALAIFLQACPAVEELREHELEIARAKKEGSELMNDYKKMDPTQ 368
Db      273 IFAADTVAVSMILKTYLAENPNVLEAVTEBQMAI-RDKREGES-LTWGDTKKMPLTS 330
QY      369 CVINTELGNVYRFLRKALKDVRKGYDIPSGKYLPIVSAVHLNDSRYDQPNLFPW 428
Db      331 RVIOETLVASILSTFEEAVEDEVEEYLIPKGRKVLPIFNHHSADIFSNPKFDS 390
QY      429 RMOOONNAGSSSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHLVLK 483
Db      391 RFE-----VAPKRTFMPFNGTSHSCGNELAKLEMIMIHLLTTY 432

```

RESULT 8

```

T02739
probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: Z14710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <ROU>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umeyam, T.; Tall,
euss, D.; Merman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: D84692
A:Molecule type: DNA
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE002093; NID:g3461849; PID:NAC33235.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:291-453/Domain: cytochrome P450 homology <P45>

```

```

Query Match      22.8%; Score 611; DB 2; Length 482;
Best Local Similarity 29.8%; Pred. No. 4.5e-32;
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;
QY      16 PSLSLFLILK-----RRNRKTRNLPKSGMPLGERTIGLYKPTATTLGDFMO 69

```

Db 20 PALITLIVVVVLLFKWMLHMKRQLRPLPGSMGLPIGET---LRLYTEP--NSFEA 75
 QY 70 OHVSKYKTYRSNLFGEPTIVSADAGLNFIITQNERLRECEYPSISGILGKMSLVLY 129
 Db 76 TRQNKMGDILFKHIILGCPVMTSSPEAAKMLVLSKAHLKPPYPSKEMIGPEALFFHQ 135
 QY 130 GDMHRDM-RSISLNLNLSHARLRTILKLDVERHTLFVLDMSQONSIFSAQDEAKKTFENM 188
 Db 136 GYHSTLKLRLVQSSFPMSALRPTV--SHIELIVLTQISWTSOKSINTLEYMKRYAFDVA 193
 QY 189 AKHISMDSGEEE---TELRKEVYTFMKGVTSAPLNPFGAYHKAQSRATILKFIERK 245
 Db 194 ---HSAGDKEEPTTIDVILKLYORLEGVNSMDLDPGLTFHKSMAKRIELSELRKY 250
 QY 246 MEERLIDKEEDOEVEEYVTEDEAEKMSKSDHYRKQRTDLDLGLWL-----KSNLSIEQ 300
 Db 251 EKRRENGREE-----GGELGVLLGAKKDKRNLSDQ 283
 QY 301 IIDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEI-ARAKKEGSEELNMD 359
 Db 284 IADNTIGVIFATDPTASVLTWMLKYLDHDPRLQDEVREQPSIQKIKKE--NRISME 341
 QY 360 DYKKMDFTQCVINETLRGNVYRELHRRKALDVRKYKGDISGKVLVYSVHLNDSRY 419
 Db 342 DTRKMLTTRVYIQLTFLRAASVLSFTFRRAVDVEYDGLYLRGKMKVLPFRIRHSSSEFF 401
 QY 420 DQPLNFPMWQOONNGASSGSGSFTWGNMYMFGGPRLCAGSELAKEMAVFIHHL 479
 Db 402 PDPEKFDPSREF-----VAKRPYTMFPGVSGVSCGSELAKEMLILLHL 448
 QY 480 VLKFMWELAEDDOPFAF-PFYDFRNGLPYRSRI 512
 Db 449 TTSFRWEYIGDEGIQIGPPVPRKKGLPIRVTP 482

RESULT 9

Cytochrome P450 d13695c - Arabidopsis thaliana
 N:Contains: oxidoreductase (EC 1.-.-.-)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Valley: Columbia
 C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 16-Jun-2000
 C:Accession: C71417
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 vanagh, T.; Hempel, S.; Kotter, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; Giel
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S.; Ans
 C.; Chaitwalis, N.
 A:Title: Analysis of 1.9 Mb of continuous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; M01D:9612113; PMID:9461215
 A:Accession: C71417
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-487 <BEV>
 A:Cross-references: GB:Z97338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888
 C:Genetics:
 A:Gene: d13695c
 A:Map position: 4COP9-4G3845
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
 F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.7%; Score 581; DB 1; Length 487;
 Best Local Similarity 27.3%; Pred. No. 4e-30;

Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;
 QY 17 SLTSLILFLILKRRRKTRFNLPPKSGMPLGRTIGLKRYATATLTDENQOHSKYG 76
 Db 3 SLFLVIFHWYQMKRPKTKNGKLPESMGSPFIFGEFFKPHDALQSFSTFKDRLRF 62
 QY 77 -----KIYRSNLGEPPTIVSADAGLNRIILQNEGRLEFECSPRSIGI-----LG 121

Db 63 ADFSSIHLSFFRTSLFGDKALISMDMEIN-----LEKANKNSPGVTKSVIRLFG 112
 QY 122 KMSMLVIVODMHRDRSRISLNLNLSHARLRTILKLDVERHTLFVLDMSQONSIFSAQDEAK 181
 Db 113 ENNLFLQKESHKVRNLTFFOLGPOGLKSRNIEVDLARTYMEGARNGYLDVKETSS 172
 QY 182 KFTFNLMAKHIM-SMDPEEETDOLKKEVYFMKGVTSAPLNPETAHKL-----232
 Db 173 KILICLAKKTVGEMEP--EAKELALCKWRFOSGWFFFLNLPETGVYKMKVLYFOYT 230
 QY 223 -----QSRATILKFIERKMEERKLDKEEDOEVEEYVTEDEAEKMSKSDHYRKQRTDLD 286
 Db 231 EADISWQARKMKMLKRLKTVLTKRSGEELGEFFNIIRGEMGE-----274
 QY 287 LGWVLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIAR 346
 Db 275 -----GETMVEANAVETITFFLVANETTRILAAVYKFTSDHPKYQELORHEETVR 328
 QY 347 AKKEGSEELNMDYKKMDFTQCVINETLRGNVYRELHRRKALDVRKYKGDIPSGMKVL 406
 Db 329 GKAE-KREGLTWEDYKSMHFTQVYINSLRITSTAPYVRLVLEHDPQVDTTIPAGTFM 387
 QY 407 PVISAVHLNDSRYDQPLNFPMWQOONNGASSGSGSFTWGNMYMFGGPRLCAGSE 466
 Db 388 G-YHIFHNSKEDYDPAFNPWRWEGDLGAIYS-----KTFIPFGAGRLLCVGAE 437
 QY 467 LAKLEMAVFIHHLVLKFMWELAEDDOPFAFPEYDFRNGLPYRS 510
 Db 438 FAKQMAVFIHHL-FRYRMKSGSTTIIRSFMLMFPGCDVOIS 480

RESULT 10

Probable cytochrome P450 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence-revision 02-Feb-2001 #text-change 23-Mar-2001
 C:Accession: A84859
 R:Lin, X.; Kail, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, G.Y
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Talldo,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; M01D:20083487; PMID:10617197
 A:Accession: A84859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <STO>
 A:Cross-references: GB:AE002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42850
 A:Map position: 2
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 579.5; DB 2; Length 485;
 Best Local Similarity 28.8%; Pred. No. 5e-30;

Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;
 QY 12 LLLSLSLILFL-----ILKRRNRKTRFNLPPKSGMPLGRTIGLKRYAT-7L 64
 Db 14 LCIATFISSTLFFPKKHNRFITKIQKKK--LLEBEGDLWIDETWDFYAKSRNRY 71
 QY 65 GPFMOHWSKYKTYRSNLFGEPTIVSADAGLNRIILQNEGRLEFECSPRSIGILGKMS 124
 Db 72 EDFVNPRIKGNINIKRTIMGSPITVVGAEANRLILNERSLVYSSMPSSSVQDLMGNC 131
 QY 125 MLVVGDMHRDRSRISLNLNLSHARLRTILKLDVERHTLFVLDMSQONSIFSAQDEAKFT 184
 Db 132 IMAGGEKHRLVLRGIVANSLSYIGLSLIPKLCDTVFKHNETEMKKEEISLYRSAYLT 191
 QY 185 FNLAAKHI--MSMDPGEETDOLKKEVYFMKGVTSAPLNPETAHKAQSRATILKFI 242


```
Db 192 FTVEECILYGIKVELGMLV-----FERVLEGVFALVEPPCSFAAKAKARLEIEFFL 245
QY 243 ERKMEERKLDIKEDOEERKEVTEDEAMSKSDHAKORTDDLLGWLKLKSNLSTEOIL 302
Db 246 VGKVRKREKREMEKGAE-----KPNITLFSRLYBELKGV-----ITEEVV 287
QY 303 DLILSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIARAKKELGSE-LNMODY 361
Db 288 DNMVLVFAAHDITSYAMSMTEFKMLAQHPTCRDITLQEHAOI---KANKGGEYLVEEDV 344
QY 362 KKMDFOCVINEITRLGNVRFELHRAKLDVRYKGYDIPSGMKVLPVSAVHLNDSRDDQ 421
Db 345 KKMYSYMQVAVETKRLSPPIGSRKKAAYADIDYGITLPKGMKILMTTYGTHNPEIFOD 404
QY 422 PNLNPMRMOQONNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVETIHLVL 481
Db 405 PMSDFPTRFDPKPIQAYT-----YLPFGGGRILCAGHQLAMISLIVENHFVYT 451
QY 482 KFNELALDDQRPAPFPVDFPN-GLPIRVS 510
Db 452 GFDMSLVYPDETISMDFLPFPISLQMPKIS 481
```

RESULT 11

hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence,revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86185
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C.A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:92388581; PIDN:ABE71462.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

```
Query Match 21.6%; Score 579.5; DB 2; Length 490;  
Best Local Similarity 28.6%; Pred. No. 5e-30;  
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;  
QY 34 KTRPNLPKSGMPELIGTYLKYATATLGDPMOQHVSKYK--IYRSNLFCEPTIYS 91  
Db 40 ENRHILPDPGLGMPFIENMLSLFAKFTSDPDSTRLIKRGKGIYKAMFGPSTIV 99  
QY 92 ADAGLNPLIONEGRLEFECYSYRISGILGKWSMLVVGDMHDMRSISLNFSLHARLT 151  
Db 100 TTSOTCARVLTTDD-APKPMPTSTMELIGKRSFVGSFEHKKRLRLTARVNGHELS 158  
QY 152 ILKDVRRHILFVLDWMQONSIFSADQAKKFTNLMAKHIMSMDPGEETDQKKEYVT 211  
Db 159 TYPIREENVITVLDKTKMGEFEFLTHLRKLFRII-MYFLSSESNVADALEREYTA 217  
QY 212 FMGVASAPNLPGTAHKALOSRAITLIKFERMEERKLDIKEDOEERKEDDEKEM 271  
Db 218 LNVGVRAAVNIPEFAHRAKAKKTLVAAPOSTVTER-----NORKNIISSKKDML 271  
QY 272 SKSDVKKORTDLDLGLWVKLHNSLSTEOILDLISLIFAGHETSSVAIALAIFLQACP 331  
Db 272 DNLNLNKK-----GKTLDEDEILIDVILMLNMGHSSGHTIMATVYFLOEHP 320
```

```
QY 332 KAVEELREHLEIARAKKELGSELMNDYKKMDFQCYINETLRLGNVRFELHRAKLD 391  
Db 321 EYLQRAKARQEMILKSRPE-GQKGLSLKETRKMEFLSYQVDETLRVITTSLSAFREAKTD 379  
QY 392 VRYKGYDIPSGMKVLPVISAVIDNSRYDQPNLFNWRMOQONNGASSSGSSTWGN 451  
Db 380 VEMNGGLIPKRGKVLVFRDVIHIDPEVFPDPKRFDPAR---DNG-----FVPRAGA 428  
QY 452 YMPFGGPRLCAGSELAKLEMAVETIHLVLKFNWELALDDQRPAPF 496  
Db 429 FLPRGAGSHLCPGNDLAKLEISIFLHFLFKLYQVAKSNBCEPVMY 473
```

RESULT 12

cytochrome P450-like protein - Arabidopsis thaliana
N:Alternate names: protein fl4d17.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence,revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T48973
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, A.; Reference number: 225008
A:Accession: T48973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <JOR>
A:Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:FL4D17.40
A:Experimental source: cultured Columbia; BAC clone fl4d17
C:Genetics:
A:Gene: ATSP:FL4D17.40
A:Map position: 3
A:Intons: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

```
Query Match 21.5%; Score 577; DB 2; Length 455;  
Best Local Similarity 28.9%; Pred. No. 6.6e-30;  
Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;  
QY 32 NKRTPNLPKSGMPELIGTYLKYATATLGDPMOQHVSKYKITSNLFCEPTIYS 91  
Db 28 NPKSNKLPKSGMGPPIIGETIDFRFYEFYELSPILKKMLRYGSLFNTILGVTYS 87  
QY 92 ADAGLNPLIONEGRLEFECYSYRISGILGKWSMLVVGDMHDMRSISLNFSLHARLT 151  
Db 88 TDKVNMETLROENKSFILSYDGLMKPLGKDSLFLKIGNIKHKKQITLHLLSSBGLR 147  
QY 152 ILKDVRRHILFVLDWMQONSIFSADQAKKFTNLMAKHIMSMDPGEETDQKKEYVT 211  
Db 148 KILKMDRYTREHLSKAKTGRLDVDAVSKLIIAHLPRKMS-----NLKPTQGA 198  
QY 212 FMKGV-----VSAPNLPGTAHKALOSRAITLIK--FIERMEERKLDIKE 255  
Db 199 KIMGIFKAPFTPMRTSYLISGKGLYNLM--ACREGRELKDIYTKMTSEK----- 252  
QY 256 EDQEEERKTEDEAMSKSDHAKORTDDLLGLWVKLHNSN-----LSTEOILDLISLIF 310  
Db 253 -----DDFLNTAIEESEKAGELLNENAIITLLPTLSC 284  
QY 311 AGHETSSVAIALAIFLQACPKAVEELREHLEIARAKKELGSELMNDYK-KMDFQOC 369  
Db 285 VTQDTTSKALICLAVFLLENPVVLAKKEH-EVLIESREDEGGTWEYHKKMFTN- 342  
QY 370 VINETLRLGNVRFELHRAKLDVRYKGYDIPSGMKVLPVISAVIDNSRYDQPNLFNWR 429  
Db 343 -----MKSSTIPAGVIMIIIPSYVNHDPRETYENPFEENPNR 379  
QY 430 WQOONNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVETIHLVLKFNWELAE 489  
Db 380 WEGKELRAGS-----KTFVFGVGLRGCAAGAFARLQISVFLHHLVTVYNSLHQ 429  
QY 490 DDQPPAPFPVDFPNGLPIRVS 511
```


Db 430 DCEVLKVPAAHLPGNISINISK 451

RESULT 13

B84733

probable cytochrome P450 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84733

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

ness, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84733

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-489 <STO>

A:Cross-references: GB:AE002093; NID:93831452; PIDN:AAC69934.1; GSPDB:GN00139

A:Gene: At2g32440

A:Map position: 2

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match Best Local Similarity 20.4%; Score 547.5; DB 2; Length 489; Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

Db 8 TLPLLLPLSLFLFLI--LKKRR-----KTRNLPKSGSGWFLGTTGYLK 57

Db 4 TGLILMFP-LIILGLFVLKWLKRVNWIYVSKLEKKHYLPBGDLGMPVIGNMSPFLR 62

Qy 58 PYATITLGDPMQOHVSKYK--LYRSNLEGEPTIVSADAGLNFILQNGRLFECSYPS 115

Db 63 AFRTSDPESFIQSYITRYGRTGYIKAHMGYPCVLTPTPCRYVLTDD--AFHIGWPS 121

Qy 116 IGGILKMSLVYVGMHMDMSISLNFSLHARLITLLKDYERHTLFVLDWQONSIS 175

Db 122 TMLIKRKSFGVGSFEHRLRLTSAIPVNGPDLVYIQTLEEVNTDLEKMSKGE 181

Qy 176 AODEAKKFTFNLMKHMMDPEEETDOLKEEYVFMKGVSAFLNPGTAYHAKLOS 235

Db 182 FLHLKRLTFKYI-MTIFLSESEHMDLEREYTNLNGVAKMGINLPFGFYHAKLAR 240

Qy 236 ----ATLKFIERKMEERKLDI--KEEQEEVEVTEDEAEMSKSDHVKORTDDLLGW 289

Db 241 KKLVAAPFOSITVNRNORONISNPKMDLIDLVDK-----NGRVLDD----- 286

Qy 290 VLKHSNLSLEQILDLILSLFAGHETSSVAIALAIFLOACPKAVEELREHLEIARAK 349

Db 287 -----EETIDLLMANTANGHSSGHLMAITLMQEHMILQAKKEQERT--VKK 335

Qy 350 ELGESELMNDYKKKMDFTQCVINETLRLGNVRLHRRKALDVRKYGADIPSGMKVLPVI 409

Db 336 RAPGCKLTLEKREKREYVLSOVIDELRLVITFSLAFRAKSDYQMDGIIYIKKAVLTWF 395

Qy 410 SAVHLDNSRYDQPNLFNPMWQOONNGSSSGSFSFTWGNVYMPFGGPRLCAGSELI 469

Db 396 RNVLHDPETLPDKKFDSSKME-----GYPKAGTFLPFGISHLGCGNDLAK 443

Qy 470 LEMAVTILHLYLKNFWEIAEDDOPAPFPVDFPGLP-----INVRIL 513

Db 444 LEISIFLHFLKLYKREVSNGPCVMF-----LPNRRKDNCLARITRTM 488

RESULT 14

D96813

hypothetical protein T30F21.17 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 19-Apr-2002

C:Accession: D96813

Query Match Best Local Similarity 18.6%; Score 497.5; DB 2; Length 460; Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

Db 13 LLPLSLSLFLFLILKRRNRRTRENLPPKSGWFLGTTGYLKPYATITLGDPMQOHV 72

Db 9 MLMVALVVRISHLWLRMSNPRCPKLPKSGWFLGTTGYLKPYATITLGDPMQOHV 68

Qy 73 SKYGYKYSNLEGEPTIVSADAGLNFILQNGRLFECSYPSISIGILGKMSLVYDGM 132

Db 69 IRGPLEKNTINIGSTVSTDEVDYHOFROHNTSFEELGYPDIFVKVGNLFLKEVFI 128

Qy 133 HRDMSISLNFSLHARLITLLKDYERHTLFVLDWQONSISFSAODEAKKFTFNLMAKHI 192

Db 129 HYYLQKITMQLIGSGGLQTMGNMDKATROHINSIASQSGFNVRKEVENLVAVYMPKL 188

Qy 193 MSMDEGEETDOLKEEYVFMKGVSAFLNPGTAYHAKLOSRAITLKFIERKMEERKLD 252

Db 189 ISNLRPEQSKLIDN-----LNAFNIDMFKSF-LRLSTKAVYKALKSR--- 231

Qy 253 IKEDQEEVEVTEDEAEMSKSDHVKORTDDLLGWVLAH-----SNLSTQILDLILS 307

Db 232 -----BEALQVAKDVLAMKRETERKO---EDLNLTLLELEKDGSPFGDSAINLIFL 281

Qy 308 LLFPGHETSSVAIALAIFLOACPKAVEELREHLEIARAKKELGESLMDVYK-KMDP 366

Db 282 LAFALREGTSSCTALAVKFSIKDPKVLAEKREHKAIVDKRD--KEAGVSWEEYRHMNTF 340

Qy 367 TOCVINELRLGNVRLHRRKALDVRKYGADIPSGMKVLPVI 425

Db 341 TNMVSNEVLRANTPLLEFRKAVDVEIKRY-----LLEVIHMGFG----- 382

Qy 426 NPMRWQOONNGASSSGSFSFTWGN-NYMPFGGPRLCAGSELIEMAVTILHLYLKN 484

Db 383 --WLMQ-----GKEMINGSKTFMAFGVNLCAEFSRLQMAFLHNLVAYD 429

Qy 485 WELAEEDOPAPFPVDFPGLP-----INVRIL 511

Db 430 FSNVQDSEITRSPFHOTKDLINTISO 456

RESULT 15

T02263

cytochrome P450 DWARF3 - maize

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Zea mays (maize)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: T02263

R:Winkler, R.G.; Helentjaris, T.

Plant Cell 7, 1307-1317, 1995

A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibber

A:Reference number: Z16468; MUID:96004534; PMID:7549486

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:51:48 ; Search time 18 Seconds

(without alignments)
1182.075 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681
Sequence: 1 MFETEHHTLPLILPLSLLS.....FAFPYDFPNGPLIRVSRLL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978.5	36.5	472	1 C901_ARATH	042569 arabidopsis
2	864	32.2	524	1 C90C_ARATH	09m066 arabidopsis
3	761.5	28.4	464	1 CP85_LYCES	043147 lycopersico
4	579.5	21.6	490	1 C881_ARATH	023051 arabidopsis
5	503	18.8	492	1 CP26_BRARE	P79739 brachydanio
6	491.5	18.3	519	1 C881_MAIZE	043246 zea mays (m
7	479.5	17.9	444	1 C120_SYNY3	059990 synecocyst
8	461	17.2	497	1 CP26_MOUSE	055127 mus musculu
9	446	16.6	512	1 C26A_HUMAN	09m163 homo sapien
10	444.5	16.6	492	1 CP26_XENLA	09p323 xenopus lae
11	439.5	16.4	492	1 CP26_CHICK	09p323 xenopus lae
12	438	16.3	497	1 CP26_HUMAN	043174 homo sapien
13	324.5	12.1	520	1 YRVA_CAEEL	027514 caenorhabdi
14	321.5	12.0	508	1 YRVA_CAEEL	027514 caenorhabdi
15	316	11.8	500	1 C912_ARATH	065790 arabidopsis
16	312.5	11.7	504	1 CP32_RAT	P05183 rattus norv
17	312	11.6	501	1 CP35_MOUSE	054749 mus musculu
18	310.5	11.6	520	1 YRVA_CAEEL	027513 caenorhabdi
19	303	11.3	501	1 CP26_MOUSE	054750 mus musculu
20	302	11.3	494	1 CPA8_MESAU	P24454 mesocricetu
21	299.5	11.2	492	1 CP8C_RAT	P33772 rattus norv
22	297	11.1	491	1 CPB9_MOUSE	P12790 mus musculu
23	294.5	11.0	491	1 CPB4_RABIT	P00178 oryctolagus
24	294	11.0	506	1 C892_ARATH	042602 arabidopsis
25	293	10.9	491	1 CPB1_RAT	P00176 rattus norv
26	291.5	10.9	530	1 CP51_YEAST	P10614 saccharomyc
27	290.5	10.8	502	1 C83A_YEAST	P48621 arabidopsis
28	289	10.8	503	1 CP39_RAT	P55538 rattus norv
29	288.5	10.8	492	1 CPB1_MOUSE	055071 mus musculu
30	288.5	10.8	503	1 CP33_HUMAN	P05184 homo sapien
31	288.5	10.7	538	1 CP18_DROME	095078 drosophila
32	288	10.7	459	1 CPXN_ANASP	P23980 anabaena sp
33	287.5	10.7	520	1 YRVA_CAEEL	027517 caenorhabdi

34	286.5	10.7	533	1 CP51_CANGA	P50859 candida gla
35	285.5	10.6	502	1 CPJ2_HUMAN	P51589 homo sapien
36	284.5	10.6	505	1 C762_SOIME	P37122 solanum mel
37	282	10.5	514	1 C941_VICIA	081117 vicia sativ
38	281.5	10.5	499	1 C831_ARATH	065782 arabidopsis
39	281.5	10.5	504	1 CP3G_MOUSE	064481 mus musculu
40	281	10.5	528	1 CP51_CANTR	P14263 candida tro
41	280.5	10.5	504	1 CP3B_MOUSE	064459 mus musculu
42	280	10.4	499	1 C771_SOIME	P37123 solanum mel
43	280	10.4	500	1 CPJ1_RABIT	P52786 oryctolagus
44	279.5	10.4	513	1 C773_SOYBN	048928 glycine max
45	279.5	10.4	519	1 YRVA_CAEEL	027520 caenorhabdi

ALIGNMENTS

RESULT 1
ID C901_ARATH STANDARD; PRT; 472 AA.
AC 042569;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 90A1 (EC 1.14.-.-).
GN CYP90A1 OR CYP90 OR CPD OR AT5G05690 OR MJJ3.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96200769; PubMed=8612270;
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
Altman T., Redei G.F., Nagy F., Schell J., Koncz C.,
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
controlling cell elongation and de-etiolation in Arabidopsis.";
RL Cell 85:171-182(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukui M.,
MYAJIMA N., Tada S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
features of the 1.6 Mb regions covered by twenty physically assigned
P1 clones.";
RL DNA Res. 4:215-230(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
or send an email to license@sdb.ch).
CC
DR EMBL; X87367; CAA60793.1; -
DR EMBL; X87368; CAA60794.1; -
DR EMBL; AB005237; BAB09663.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21 POTENTIAL.
FT BINDING 418 418 HEME (BY SIMILARITY).
SQ SEQUENCE 472 AA; 53785 MW; 41A73F46D64E343F CRC64;
Query Match 36.5%; Score 978.5; DB 1; Length 472;
Best Local Similarity 41.9%; Pred. No. 1.7e-51;

Matches 210: Conservative 79: Mismatches 171: Indels 41: Gaps 8:

QY 12 LLLPSLLSLFLLLKRRNRKRNLPKSGMPGEGTIGYKPYATTLLGDFMOOH 71
 Db 7 LLLSSISAGFL---LLRRTYRRMGMPLPGSLGPIGLGEPQLGAYKTEPEPFI 63
 QY 72 VSKYKIRSNMFEPTIVSADAGNRFLLONEGLFCGYSRISGILGKSMIVGD 131
 Db 64 VARGSVMTLFGPEPTIFSADPEPNRYVLONEGLFCGYSRISGILGKSMIVGD 123
 QY 132 MHRDRSISLNLFLSHARLRTLLKDNERTFLVDSMOONSIFSAODEAKKTFNLM 191
 Db 124 LKRNHSLTMSFANSIIKDLMLDIDRLVRNLDSSWSRYLL--MEKAKTTFELTY 181
 QY 192 IMSNDPGEETDOLKEVYTMKGVVAPLIPGAYIKALOSRATILKFERKKEER 251
 Db 182 LMSDPPG-EMSESLRKEFLVIEGFFSLPLPFFSTYRKAIQAR-----RK 232
 QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGVWLKHSNLTSTOII 311
 Db 233 VVWAKRREBEGBE-----RKMDLALLAADGFSDEIDELVALLVA 278
 QY 312 GHETSSVAIALAIFFLQCPRAVELREHLEIRAKKELGESELMMDYKMDTC 371
 Db 279 GYETSTIMTLAVKFLFTPLALQKHEHEKIRAKMD--SYLEMDSKSMPTOC 336
 QY 372 NETIRLGNVVRFLRKALKDVRIGYDIPSCMKVLPVISAHLDSRYDQPL 431
 Db 337 NETIRVANIIGGVRRAMTDVEIKGYKIPKMGVFSSRAVHLDPNHFED 396
 QY 432 QONNGASSGSGSSTGMN--YMPGGPRCAGSELAKLMVAVIHLVLKFNEL 490
 Db 397 -----SNSTTTPGSNVTFFPGGPRCPGIELARVALSVFLHRLVGF 446
 QY 491 DQFAFPFVDFPNCPIRVS 511
 Db 447 DKLVEFPTRTKRYPIFKR 467

RESULT 2
 C90C_ARATH STANDARD: PRT: 524 AA.

AC Q9M066: 023242;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 41, Last annotation update)
 DE Cytochrome P450 90C1 (EC 1.14.-.-) (ROTUNDIFOLIA3).
 GN ROT3 OR CYP90C1 OR AT4G36380 OR C7A10.980 OR F23E13.220.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eucosids II: Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98361880; PubMed=9694802;
 RA Kim G.-T., Tsukaya H., Uchimiya H.,
 RT "The ROTUNDIFOLIA3 gene of Arabidopsis thaliana encodes a new member
 of the cytochrome P-450 family that is required for the regulated
 polar elongation of leaf cells."
 RL Genes Dev. 12:2381-2391(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Columbia;
 MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wandutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delisny M., Pulgomech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohsels J., Zimmermann W., Weller H., Ridley P.,
 RA Langham S.-A., McCullagh B., Blham L., Robben J.,

RA Van der Schuren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Wellens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Willemsen T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirke P.,
 RA Moellman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail A., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Pettit A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Cabel C., Fuchs M., Fartmann B., Grandenath K., Dauner D., Hertzl A.,
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravalli E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlin A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedid F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Jaco D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Fishman D., Haase D., Lemcke R., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parrell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kallick J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Speth J., Ryan E., Andrews S., Giesel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamer B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Rodriguez M., Hoffman J., Tili S.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana."
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE BIOSYNTHESIS OF STEROIDS
 CC IMPORTANT FOR THE POLAR ELONGATION OF CELLS DURING DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- TISSUE SPECIFICITY: UBQUITOUS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB008097; BA37167.1; -
 DR EMBL: AL161589; CAB80304.1; ALT_INIT.
 DR EMBL: Z99708; CAB16850.1; ALT_INIT.
 DR EMBL: AL022141; CAA18139.1; ALT_SEQ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
 KW Endoplasmic reticulum; Multigene family.
 FT TRANSMEM 4
 FT BINDING 463 463 HEME (BY SIMILARITY).
 FT CONFLICT 45 45 F -> L (IN REF. 2).
 SQ SEQUENCE 524 AA; 59389 MW; 550578908BDDP272 CRC64;

Query Match 32.2%; Score 864; DB 1; Length 524;
 Best Local Similarity 35.4%; Pred. NO. 1.3e-44;
 Matches 186; Conservative 96; Mismatches 165; Indels 78; Gaps 10;

24 FLIL-----LKRNRKRN-----LPPGSGMPFLGTYL---

```

Db 30 FLVLAGILLRPMWFLRNRKSKTKDGEDEENEEKKMGKMPNGSLGMPVIGETLNFAC 89
QY 57 ---KPTATVLDGDMQOHVSKYKIRSNLFGEPITVSADAGNRIPLONEGLEFECSTP 113
Db 90 YSRPT-----FMKRSLSLGKVRKPTNIGTPILISDAENKVLQMHGNTFPAYP 143
QY 114 RSIGGLGKMSMLVLVGDHMRDRSISLNFLSHARLTLLKDYERHTEFLDSMOONS 173
Db 144 KSITTELGNSLISINGPQKRLHTLIGAFRLSPHLKDRTRDIEASVITLASMOLPL 203
QY 174 FSAODEAKFTFNLAHAKHIMSMDPGEETDQKKEYVTMKGVASAPLPGTAHKAIO 233
Db 204 VHVODEIKMTFEIIVKVLMSSTPG-EDNNILKLEFEETIKGLICIPKPGTRLYSK 262
QY 234 SRATILKFERKMEERKLDIKEEDQEEVEK-----DEAMSMGSDHVRKORTDOLL 287
Db 263 AERLIKMYKAYVEEQVAMTTSPANDVAVDLLKDGSGSEKOPSDFVSGK----- 315
QY 288 GMYLHNSLSTEOILDLISLFGHETSSVAIALAIFLQACPKAVEELREHELTARA 347
Db 316 -----IYEMIPGETMPTAMTLAVKFLSDNPVALAKIVEENMEKKR 358
QY 348 KEELGESELMDDYKMKMETQCYINETLRLGNVRFLLHRAKALKDVRKYGYDIPSGKVL 407
Db 359 KLELGE-EYKWDYMSLFTQVYNETLTMANINGVMRKALKDVEIKGLIPKGCYLA 417
QY 408 VLSAVHLNRSRYDOPNLFNPMRMOQONNGASSSGSFTGWNKNNPFGGPRLCAGSEL 467
Db 418 STSYHMDIDYDNPQFDPMDRDRINGSANSSIC-----FPPGGGRLCGLEL 468
QY 468 AKLEMAVFLHLLVLFKNMELADDOFAPFPVDFPGLPIRVRSRI 512
Db 469 SKLEISIFLHLVTRYSW-TAEDEIVSFPTVKMKRRLPIRVATV 512

RESULT 3
CP85_LYCES
ID CP85_LYCES STANDARD: PRT: 464 AA.
AC 043147;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 85 (EC 1.14.-.-) (Dwarf protein).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=9626705; PubMed=8672892;
RA Bishop G.J., Harrison K., Jones J.J.G.D.;
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
RT encodes the first member of a new cytochrome P450 family.";
RT Plant Cell 8:959-969(1996).
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U54770; AAB17070.1;
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.

```

```

KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BYNDING 414
SO SEQUENCE 464 AA; 53706 MW; DZB21AAB7B14E94 CRC64;
Query Match 28.4%; Score 761.5; DB 1; Length 464;
Best Local Similarity 33.4%; Pred. No. 1,4e-38;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

QY 12 LLLPSSLILLFLILKRN--RKTFNLPFGKSGMPETIGLYKPTATYLG-DEM 68
Db 5 LIFLSFFGLICIFCTALLRWNQVKNKNPPTMGMPGLGETTEFLK-----LGPSFM 58
QY 69 QOHVSKGYKTYRSNLFGEPTIVSADAGNRIPLONEGRLECSPRSIGGLGKMSMLV 128
Db 59 KNQARIGSFYKSHIIGCPTIVSMDSLNRIYLNENKGLVPGTPQSMIDLIGCNTAAV 118
QY 129 VGDHMRDRSISLNFLSHARLTLLKDYERHTEFLDSMOONSIFSODEAKFTFNLM 188
Db 119 NSAHKMYRGAALSLISPTMIRQQLPKIDEPKRSHTLNM-DNKVIDIQETKNMFLSS 177
QY 189 AKHIMSMDPGEET---EQLKKEYVTMKGVASAPLPGTAHKAIOSRATYIKTERK 245
Db 178 LKQI---AGIESTSLAQEFMSSEFNVLGTLSPMLPMTNVRHGFQARKIYVNLRTL 233
QY 246 MEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKORTDOLLGMYLHNS-----NLSTEOI 301
Db 234 IEERR-----ASKELQMDLGLTAMNEATREKLTDEM 266
QY 302 LDLISLFGHETSSVAIALAIFLQACPKAVEELREHELTARAKEELGESELMDDY 361
Db 267 IDLITILYSGYEIVSTSMVAVKYLLHDHFKVLEELKEHMAIREKKP--EDPIDYNDY 324
QY 362 KKMDFOQVINEFLRLGNVRFLLHRAKALKDVRKYGYDIPSGKVLPIVSAVHLNRSRYD 421
Db 325 RSMRFTVAVTLERSLATIYNGVLRKTTQDMEINGYIIPGWRYYVYTRRLNDPRLYPD 384
QY 422 PNLFPNPMOQONNGASSSGSFTGWNKNNPFGGPRLCAGSELAKLEMAVFLHLLV 481
Db 385 PYSFNPMRWMDKS-----LEHQNSEFLVFGGTRGCPKRELGVATISFTLHFTV 433
QY 482 KENMELADDOFAPFPVDFPGLPIRVRS 510
Db 434 KYRMEIIGDKLKMFPVEAPNGLIRIVS 462

RESULT 4
C883_ARATH
ID C883_ARATH STANDARD: PRT: 490 AA.
AC 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 88A3 (EC 1.14.-.-).
GN CYP88A3 OR AT1G05160 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Arujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elugu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gali J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzler L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenská I., Kuritz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

```

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pei G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:816-820(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AC000098; AAB71462.1; -
 DR HSSP: P14779; 1JPZ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
 FT TRANSMEM 6
 FT BINDING 439 439 HEME (BY SIMILARITY).
 SQ SEQUENCE 490 AA; 56409 MW; 7ED8CD7A8864D298 CRC64;
 Query Match 21.6%; Score 579.5; DB 1; Length 490;
 Best Local Similarity 28.6%; Pred. No. 1,1e-27;
 Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
 QY 34 KTRFNLPPGSGKPELGETIGYLYKPYATATLGDPMQOHVSKYK--TYRSNLFGEPTIVS 91
 DB 40 ENHYHLPDGLGMPFIGNMISFLRAFTSDPDSTRLLIKRYGKGIYKAMFENPSIIV 99
 QY 92 ADAGLNFILIONEGRLFCESYPSNIGSILGKWSMLVLVGDHMDRSISLNFSLHARLRT 151
 DB 100 TTSDTGCRFVYLDLDD-AFKPCMPSTMLIGRSKSEVSGISFEHKKRLRLTAPVNGHEALS 158
 QY 152 ILKLDYERHFLVLDNQNSISFSAODEAKKFTFNLMAKHMSMDPDEEIEQAKKEYVT 211
 DB 159 TYIPYEENVTYLDKTKGEEFEFTLHKKLTFRII-MYIFLSESENVADALEREYTA 217
 QY 212 FMGVSAPLPLPGTAVYHAKLQSRATILKFERKMEERKIDKEDEEVEKTEDEAEM 271
 DB 218 LNVGVRAAMVNTGFAVHRLAKRTIYVAFQSIYVTRR-----NORKNIIISNKKDML 271
 QY 272 SKSDHVRKQRTDDLLGWLKHSNSTEQIIDLILSLFAGHETSVAVALAIFFLQACP 331
 DB 272 DNLINLVK---DEB-----GKTLWDEETIDVLMYLNAGHESGHTIMAVYFLDHP 320
 QY 332 KAVEELREHLELAKAKKEISELNDVKKMDFTQCYNETRLGNGVAFRLHKKALKD 391
 DB 321 EVYQRAKAEQEMILKSRPE-GQKGLSLKTERKMEFISQVAVDETALRTVTFSLTAERAKTD 379
 QY 392 VRYKGDIPSGKVVLPVYSAVHLDNSRYDQNLFNPMRMOONNGASSGSGSFTWGN 451
 DB 380 VEMNGVILPKGMVLTWFRVYHIDPEVFRPKREDPRRW---DNG-----FVPRAGA 428
 QY 452 YMFPGGPRICAGSELAKLEMAVFIHHLVLKFMWELADDDQPEAF 496
 DB 429 FLPEFGAGSHLCPGNDLAKLEISIFLHFLKLYQVKRSNPECPVAY 473

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
 DE cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).
 GN CYP26.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97094702; PubMed=8939936;
 RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
 RA Hsu K.E., Dilworth F.J., Jones G., Petkovich M.;
 RT "Identification of the retinoic acid-inducible all-trans-retinoic
 RT acid 4-hydroxylase".
 RL J. Biol. Chem. 271:29922-29927(1996).
 CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
 CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
 CC STEROISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION; HYDROXYLATION.
 CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
 CC INCLUDING 4-OH-RA AND 4-OXO-RA.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- INDUCTION: BY RETINOIC ACID.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U68234; AAC60045.1; -
 DR ZFIN: ZDB-GENE-980415-44; Cyp26.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum.
 FT BINDING 438 438 HEME (BY SIMILARITY).
 SQ SEQUENCE 492 AA; 56281 MW; F0471435B2F30509 CRC64;
 Query Match 18.8%; Score 503; DB 1; Length 492;
 Best Local Similarity 27.6%; Pred. No. 3.8e-23;
 Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;
 QY 7 HTLLPLPLSLLSLLFL-----ILKRRNRKTRFNLPPGSGKPELGETIGYLYK 58
 DB 4 YTLWTFELCTIVLPVLLFLAVALKLEMLMIRVDPNCRSPPLPGTMGLPFIETQLI-- 61
 QY 59 YTATLIDPMQOHVSKYKITSNLFGEPTIVSADAGLNRFILQNGRDFECESYPSISIG 118
 DB 62 ---LQRKPELKRKKRYKCYITKTHLFGNVPYKMGADNVKQLDLAEHKLVSQVMPASVYT 118
 QY 119 ILGKWSMLVLVGDHMD-----MRSISINFLSHARLRTILKDYERHFLVLDNQNS 172
 DB 119 ILGSDTILSNVAGVGHKKKKKIMRAFSDALEH-----YIPVIOQVKSALQENLQNS 172
 QY 173 IFSADDEAKKFTFNLMAKHMSMDGEET--EOLKKEYVTFMKGVSAPLPLPGTAVYK 230
 DB 173 CVLVPEPKKMLFRIAMILLGFEEQIKTDEQELVEAFEEIKMLFSLPDIVPESGLYR 232
 QY 231 ALQSRATILKFERKMEERKIDKEDEEVEKTEDEAEMSKSDHVRKQRTDDLLGMY 290
 DB 233 GLRAR---NFIHSIEENIKKKIQQDDNENQKTKDLQLD---IENSRSDE----- 279
 QY 291 LKHSNLSLEQIIDLILSLFAGHETSVAVALAIFFLQACPVAEELREHLEIARAKE 350
 DB 280 ----PFLQAMKKEATLFLFGCHETTAATAISIVAFGLNTEVYQKVAEE----VQEKVE 331

```

OY 351 LG-----ESLNDNDYKRMDFQCVINETLRNGVNFELHRRKALDKVRYKGYDIPSGMKVL 406
DB 332 KMATYRPGKLSMELLDQLKYTCGVICKEITLRINPVPGRFVALKTEFLNGLYQPKGMNVI 391
OY 407 PVIASVHLDNSRYDQNLNPNRMWQOONNGASSSGSGSPSTGMNNMPECGGRLCAGSE 466
DB 392 YSICDTHADVADVPNKEEFQPERFM-----SKGEDGSRF--NYIPFGSGRMCGVKE 442
OY 467 LAKLEMAVFIHHLVLFKFNWLEADDQ-----PPAFVDPDP 502
DB 443 FAKVLKLTIVELTQHCNWLISNGPMTKGTPIITYVDNLP 483

RESULT 6
C881_MAIZE
ID C881_MAIZE STANDARD: PRT: 519 AA.
AC 043246:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 88A1 (EC 1.14.-.-) (DWARF3 protein).
GN CYP88A1 OR D3.
OS Zea mays (maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OC NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. B73.
RA MEDLINE=96004534; PubMed=7549486;
RW Winkler R.G., Helentjaris T.;
RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
RL in gibberellin biosynthesis.";
RT Plant Cell 7:1307-1317(1995).
CC -1- PATHWAY: gibberellin biosynthesis; early step.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32579; AAC49067.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 1 21
FT BINDING 466 466 HEME (BY SIMILARITY).
SQ SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match 18.3%; Score 491.5; DB 1; Length 519;
Best Local Similarity 28.2%; Pred. No. 2e-22;
Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

OY 36 RFNLPFGKSGWPLFTIGYIKRYPTATTTGDFMQOYKSYK--ITRSNLFGEPTI-VSA 92
DB 69 RARLPFGKSGWPLVGVGMALFRAFKGKDPAFASVVRGRGTVRSFESSPTVYVTT 128
OY 93 DAGLNFELIQNGRLEFECYSPNSIGILKWSMLVLVGMHMRMSISLNFSLHARLRTI 152
DB 129 ASGCKGVLMDDDA--FVTGMPKATVALVPRSEFVAMPYDEHRRIRKLTAPINFALTLG 186
OY 153 LAKDVERHLLFVLDSW-QQNSIFSADDEAKKFTFNMAHIMNDGCEETEDLQKEYVT 211
DB 187 YLPFTDRIVTSSLRAMADHGSVEFTLRLRMTEFKTIIVQ-IFLGADQATRALERSYTE 245

```

```

OY 212 FMKGVSAPLNPCTAYAHKALQSRATILKFIERRMEERKLDIKE-----EDOEEREYKT 265
DB 246 LNYGMRMAINLPFAFARGLARRRRLVAVLQGLDERARARAGVSGGVDMDRLIEA 305
OY 266 EDEAEKMSKSDHVRKQRTDLDLGVNLKHSNLSTQILDLLISLPAGHETSSVAIALAIF 325
DB 306 QDE-----KGRHLDLDD-----EILDVLMYNLAGHSSGHTMMATV 342
OY 326 FLQACPKAVELREHEHLIARAKKELGESELNMDQYKMDTQCVINETLRNGVNFELH 385
DB 343 FLQENPMPFARAKQDEALMKSIPS-SQKGLTLDFKRMETLSOVIDETLRLVINSFVSF 401
OY 386 RKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLNPNRMWQOONNGASSSGSGSE 445
DB 402 RQATRDVYFVNGYLLPKMKQVGLWVRSHVMDQVYPPDPKFPSPSRE-----GHSRACTF 456
OY 446 STWGNNTMPPGGGPRCLCAGSESLAKLEMAVFIHHLVLFK 483
DB 457 -----LARGTARLCPGNDLAKLEISVFLHFLGLGY 487

RESULT 7
C120_STYNY3
ID C120_STYNY3 STANDARD: PRT: 444 AA.
AC 059990:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative cytochrome P450 120 (EC 1.14.-.-).
GN CYP120 OR CYP OR SLR0574.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OC NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64003; BAA10496.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
FT BINDING 391 391 HEME (BY SIMILARITY).
SQ SEQUENCE 444 AA; 50578 MW; 8F62A9EED3B54BDC CRC64;

```



```

Db      336 LCKSNQDNKLDMEFLLEQLKXIGVYKETTLLRNPVPGFGRALAKTFELNGYQJPKGNVI 395
QY      407 PVISAVALHDSRYDQPLFNPWRMQOONNASSGSGSFSTWGNNTYPPFGGRLCAGSE 466
Db      396 YSICDTHDVADIFTNKEEFPDRFIVHPEDASFS-----FIPFGGRLSCVCKE 446
QY      467 LAKLEMAVFTHLVLKFNWELAE-----DDQPPAFPEVDP 502
Db      447 FAKLLKIFVELARHODQLNGLPPTMKTSPVYVDNLP 487

RESULT 9
C26A_HUMAN
ID C26A_HUMAN STANDARD; PRT; 512 AA.
AC 09NR63; 09NP41;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26A2 (EC 1.14.--) (P450RA1-2) (Retinoic-acid
  metabolizing cytochrome).
GN CYP26A2 OR P450RA12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taint M., Stangle W., Zhang A., Everingham S.,
RT "Identification of the human cytochrome P450, P450RA1-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism."
RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Pape K., Jones T.;
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Involved
CC in the specific inactivation of all-trans-RA. Responsible for
CC generation of several hydroxylated forms of RA, including 4-OH-RA,
CC 4-oxo-RA, and 18-OH-RA.
CC -1- ENZYME REGULATION: Has a preferred activity towards the following
CC substrates: all-trans-RA > 9-cis-RA > 13-cis-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, particularly in the
CC cerebellum and pons.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF252297; AAF76003.1; -
CC EMBL: AC007002; AAF55576.1; -
CC MIM: 605207; -
CC
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450_2.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450_1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
CC Endoplasmic reticulum
CC BINDING 441 441 HEME (POTENTIAL).
CC CONFLICT 68 68 Q -> QKCTRETRVWMLQ (IN REF. 2).
CC SEQUENCE 512 AA; 57512 MW; A06DID944E6726F CRC64;

```

```

Query Match 16.68; Score 446; DB 1; Length 512;
Best Local Similarity 26.98; Pred. No. 9.9e-20;
Matches 132; Conservative 86; Mismatches 219; Indels 54; Gaps 10;

QY 14 LPSLSLLFLILKLR-----NKRTRFLPPGKSGWPLGFTIGLYKATYATTL 64
Db 15 LAACLVSYTLILLASQOLRMATYDKSCRLIPFGSMGFPLIGETGMLLOGSG--- 71
QY 65 GDFMOQHVSKYKGIYRNSLGEPTIVSADAGLNRFILQNGRLEFCSPYSIGILGKMS 124
Db 72 --FOSSRREKYGVNFKHLLGRPLIRYTGAEVNNRKILMGHHLYSTWEPSTRLALPNT 129
QY 125 MLYLVGDMDHMDMSISNPLSHARLTILIKDYERHTLFYLDSDMQON-SIFSADQAKKF 183
Db 130 VSNISGDIHNKRRKRVSKIPSHALESYLKR-10LV1QDPLRAMSSHPKAIINYQEQKL 188
QY 184 TFMIMAKHIMSMDPGEETEQLKREYVTEFKGVVSPALNPGTAHYKALOSKATILKFE 243
Db 189 TFMATIRVLIGFSIPEEDLGLHEVYQGFVNVFSLPVDLPFGSGRRGIQARQILOGLE 248
QY 244 RKMEERKLDKEEDDEEVEKTEDEAKMSKDVHVKQRTDDLLGWLYKHSNSTEQILD 303
Db 249 KAIREK----LQCTGQDYDALDLILESSKEHGK-----WFMQELKD 288
QY 304 LILSLFAGHETSVAIALAIFLOACPKAVEELREBHLEIARAKKELG-----ESELN 357
Db 289 GTLELLEFAVYATTAASASTSLIMQLNHPVLEKLKDE---LRAGHILHSGGCPCEGTLR 344
QY 358 WDDYKRMDFLOCIVNFTLRIGNVRFRLRKALKADVAYKGYDPSGKVLVYSVAHLDS 417
Db 345 LDTLSLRILDCYIKRYMRLFTYISGGRVLTQTFELDFQILKGSVMSIINDTDTAP 404
QY 418 RYDQPLFNPWRMQOONNASSGSGSFSTWGNNTYPPFGGRLCAGSELAKLEMAVFTH 477
Db 405 VFQDVNVFDPDRFSQ---ARSEDKDGRF-----HYLPFGGVYTCGKHLAKFLVLAV 456
QY 478 HLYLKNWELA 488
Db 457 ELASTSRFEELA 467

RESULT 10
CP26_XENLA
ID CP26_XENLA STANDARD; PRT; 492 AA.
AC 093323;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26 (EC 1.14.--) (Retinoic acid degrading enzyme
  CYP26) (XCP26) (Retinoic acid converting enzyme) (RACE).
GN Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99077803; PubMed=9857192;
RA Hollemann T., Chen Y., Grunz H., Pieker T.;
RT "Regionalized metabolic activity establishes boundaries of retinoic
RT acid signalling."
RL EMO J. 17:7361-7372(1998).
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- INDUCTION: By retinoic acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


QY 287 LGWVLSKHS-----NISTEOLIDLILSLFAGHETSVAIALAIFFLQACRAVEELREH 341
 DB 266 LQMLEHTOGNGEDLNQMLEKESATLILFGHETASANSILIFLGI-----HH 315
 QY 342 LEIARAKEL-----GESELMNDYKKMDFTOCVINETRLGNVFLRKALKD 391
 DB 316 DVLOKVRKELOKLGKLSGPNQEROLNMFLEQLKTYGCVIKETRLSPVPGGRIALKT 375
 QY 392 VRYGVYDIPSGWKVLPYISAVHLDNSKRYDQNFNPMWQOQNGASSSGSFGSTWGN 451
 DB 376 LEINGYDIPKGMWNYISICDTHVADLFTDKDEFNPRFM-----SPSPEDSSRPS----- 426
 QY 452 YMPGGPRLCAGSELAKLEMAVFIHHLVLFKNELAEADQ-----PEAPFVDFP 502
 DB 427 FIPGGGLRSCGCVGEFAKVLKIFTVELARSCWQMLNGPPTKGTGIVPVDNLP 482

RESULT 12
 CP26_HUMAN STANDARD; PRT; 497 AA.
 ID AC 043174;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
 GN CYP26A1 OR CYP26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97373542; PubMed=9228017;
 RA White J.A., Beckelt-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
 RA Jones G., Petkovich M., van der Burg B., van der Saag P.T.;
 RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
 RT identifies a novel family of cytochromes P450.";
 RL J. Biol. Chem. 272:18538-18541(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98380037; PubMed=9716180;
 RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schalkes R.K.,
 RA Petkovich M., van der Burg B., van der Saag P.T.;
 RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
 RT all-trans-RA and can be induced through RA receptors in human breast
 RT and colon carcinoma cells.";
 RL Cell Growth Differ. 9:629-637(1998).
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=99045433; PubMed=9826557;
 RA Trofimova-Griffin M.E., Juchau M.R.;
 RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
 RT cephalic tissues.";
 RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
 CC -!- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
 CC retinoids, including all-trans-retinoic acid (RA) and its
 CC stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
 CC hydroxylation. Responsible for generation of several hydroxylated
 CC forms of RA, including 4-OH-RA, 4-oxo-RA, and 18-OH-RA.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -!- TISSUE SPECIFICITY: Highest levels in adult liver, heart,
 CC pituitary gland, adrenal gland, placenta and regions of the brain.
 CC -!- INDUCTION: By retinoic acid.
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC
 DR EMBL; AF005418; AAB88881.1; .
 DR HSSP; P14779; 1JPZ.
 DR Genew; HGNC:2603; CYP26A1.
 DR MIM; 602239; .
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Mitochondrion;
 KW Endoplasmic reticulum.
 FT BINDING 442 442 HEME (POTENTIAL).
 FT SEQUENCE 497 AA; 56162 MW; EAB684B24B2EAB3 CRC64;
 SO

Query Match 16.3%; Score 438; DB 1; Length 497;
 Best Local Similarity 27.08; Pred. No. 2.9e-19;
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLTLK-----RRNRKTRNLPKSGMFLGTTGLK 57
 DB 3 LPALLASALCTFVLPPLFLFAIKIMDLVCVSGHRSICALPLPGTGWPFPGFTL----- 58
 QY 58 PYATATLGDPMQOHVSRYGKIYSNLGEPYISADGLNRFILQNGRLFECSYPSIG 117
 DB 59 -QWVLOKRRKFLQKKRRYGTHTLFGRTYRMGADVNRILLGDRLVSVMHPASVR 117
 QY 118 GILGKSMVLVGDMDHMRMSISLNFSLSHARL---TILKDYERHTLFLVDSMQQ--NS 172
 DB 118 TILGSGCLSNLHDSHQRKVKVIMRAFSLCALCYPVITEVGGSS-----LEQMLSGER 173
 QY 173 IFSAODEAKKFTTNLAKHIMSDP-----GEETDQLKKEYVTFMKGVASADLNLPGTA 227
 DB 174 GLIYVEYVKRKLRIAMRILGCEPDLAGDSE--QOLVEAFEMTRNLFSLIDVPSG 232
 QY 228 YKALDSRAITLFIERKMERKLDIKEED-----QEEVEKTEDEAKMSHVAKQRT 282
 DB 233 LYRGKMARNLIHRIQONIRAKICGLRASAGGCCDADQLLEHSE-----GERTL 285
 QY 283 DDDLLGWVLSKHSNSTEOLIDLILSLFAGHETSVAIALAIFFLQACRAVEELREH 342
 DB 286 DMQ---ALKOS--STE-----LLFGHETASANSILTYLGIVHLQVREELK 331
 QY 343 LEIARAKELGESELMNDYKKMDFTOCVINETRLGNVFLRKALKDVRKYGIPIG 402
 DB 332 SKGLCKSMQDKLDMELIQLKTYGCVIKETRLNPPVPGGRVAKLTFELNGYQIPRG 391
 QY 403 WKVLPYISAVHLDNSKRYDQNFNPMWQOQNGASSSGSFGSTWGNVMPGGPRLC 462
 DB 392 WNYIYSICDTHVADLFTDKDEFNPRFM-----FIPGGGLRSC 442
 QY 463 AGSELAKLEMAVFIHHLVLFKNELAE-----DDOAPFVDFP 502
 DB 443 VKEERAKLILKIFTVELARSCWQMLNGPPTKGTSTVYPVDNLP 487

RESULT 13
 YRY2_CAEEL STANDARD; PRT; 520 AA.
 ID AC 027514;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative cytochrome P450 CYP13A5 (EC 1.14.-.-).
 GN CYP13A5 OR TIOB9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditidae.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Gardner A.;

RT Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; Z48717; CA88604.1; -;
 DR HSSP; P14779; IJP2.
 DR WormPep; T10B9.2; CE01655.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SQ SEQUENCE 520 AA; 59524 MW; 0B7B19E25B7ADFB3 CRC64;

 Query Match 12.18; Score 324.5; DB 1; Length 520;
 Best Local Similarity 23.68; Pred. No. 1,7e-12;
 Matches 120; Conservative 94; Mismatches 178; Indels 117; Gaps 20;

 QY 42 GKSGMPFLG-----ETIGYLRKYATATLGDPMQOHVSKYGIYSNLFGEPTIYSAD 93
 DB 33 GPRGLPFLGVIHNFODYENPGLK-----LGEWTK-----YGIYGTTECVETLTVSN 82
 QY 94 AGLNRFILQNGRLFECSYPSISGILG-----KMSMLVLY-GMHMDRISISNLFSHA 147
 DB 83 P---EFVHEVYVQDFNFGKRTNPLOGDPNKNKRALVSAQGHMKRLRLTASPTFSNK 139
 QY 148 RLRTIL-----LKDVERHTLFVLDWQONSISADDAKFTFNMAKHISMDEE 201
 DB 140 NLKRIMSTVEETVELMRH---LDDASAKKAVDLIDYQOFTLIDIGRIAM---GOTE 192
 QY 202 TEOLKREYVPMKG-----VSAPLNLPGTAVHK-----ALQSRATILIKF 241
 DB 193 SLMFRRNMLPKVIGIFKDGKRLPPLVSGIFPIAGTMRREFPMRPSIQAPVDINSTEKA 252
 QY 242 IERKMERKIDKE-----EDQ-----EEEEKTEDEAKMSKSDHYRK 279
 DB 253 LNKRLRGRADEKAGIEPSEPDIDFLDARANDVFEEBSALGKAKTEIAVD---K 309
 QY 280 QRTDDLLGVNLKHSNLSLSTQIIDLILSLFAGHETSSVAIALAIFFLQACPKAVEELRE 339
 DB 310 QLTFDELITG-----QLFVFLAGYDTALSLSSSYLLARHPEIQKLOE 354
 QY 340 EHELIARAKKELGESSELMNDYKMDFTOCVINEITLGNVYRFLH-RKALKDVRKYGYD 398
 DB 355 E-----VDEECNPEPTTFQISKLYKMECVKEALMPLASTVYNNRCKKEKENVLGVQ 408
 QY 399 IPSGKMYLPVISAHLDNSRY-DQPLFNPMRMOQONNGASSSGSFSFGNNYPMFGG 457
 DB 409 IEKGTNVQVDTWTLHDPKVMGEDANFEPRME-----SGDELFPYAKG-GYLPFGM 459
 QY 458 GPRLCAGSELAKLEMAVFIHHLVKFNWE 486
 DB 460 GPRICIGMLAMMEKKMLTHILKKYTFE 488

 RESULT 14
 ID YRYA_CAEEL STANDARD; PRT; 518 AA.
 AC Q27519;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
 GN CYP13A7 OR T10B9.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Gardner A.;
 RT Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; Z48717; CA88609.1; -;
 DR HSSP; P14779; IJP2.
 DR WormPep; T10B9.10; CE01655.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SQ SEQUENCE 518 AA; 58999 MW; CC04283BF87B9EA7 CRC64;

 Query Match 12.08; Score 321.5; DB 1; Length 518;
 Best Local Similarity 24.08; Pred. No. 2.6e-12;
 Matches 120; Conservative 96; Mismatches 184; Indels 101; Gaps 20;

 QY 42 GKSGMPFLG-----ETIGYLRKYATATLGDPMQOHVSKYGIYSNLFGEPTIYSAD 93
 DB 33 GPRGLPFLGVIHNFODYENPGLK-----LGEWTK-----YGIYGTTECVETLTVSN 82
 QY 94 AGLNRFILQNGRLFECSYPSISGILG-----KMSMLVLY-GMHMDRISISNLFSHA 147
 DB 83 P---EFVHEVYVQDFNFGKRTNPLOGDPNKNKRALVSAQGHMKRLRLTASPTFSNK 139
 QY 148 RLRTI-----LKDVER-----HTLFVLDWQONSISADDAKFTFNMAKH 192
 DB 140 NLKRIMSTVEETVELMRH---LDDASAKKAVDLIDYQOFTLIDIGRIAM---GOTE 192
 QY 202 TEOLKREYVPMKG-----VSAPLNLPGTAVHK-----ALQSRATILIKF 241
 DB 193 SLMFRRNMLPKVIGIFKDGKRLPPLVSGIFPIAGTMRREFPMRPSIQAPVDINSTEKA 252
 QY 242 IERKMERKIDKE-----EDQ-----EEEEKTEDEAKMSKSDHYRK 279
 DB 253 LNKRLRGRADEKAGIEPSEPDIDFLDARANDVFEEBSALGKAKTEIAVD---K 309
 QY 280 QRTDDLLGVNLKHSNLSLSTQIIDLILSLFAGHETSSVAIALAIFFLQACPKAVEELRE 339
 DB 310 QLTFDELITG-----QLFVFLAGYDTALSLSSSYLLARHPEIQKLOE 354
 QY 340 EHELIARAKKELGESSELMNDYKMDFTOCVINEITLGNVYRFLH-RKALKDVRKYGYD 398
 DB 355 E-----VDEECNPEPTTFQISKLYKMECVKEALMPLASTVYNNRCKKEKENVLGVQ 408
 QY 399 IPSGKMYLPVISAHLDNSRY-DQPLFNPMRMOQONNGASSSGSFSFGNNYPMFGG 457
 DB 409 IEKGTNVQVDTWTLHDPKVMGEDANFEPRME-----SGDELFPYAKG-GYLPFGM 459
 QY 458 GPRLCAGSELAKLEMAVFIHHLVKFNWE 486
 DB 460 GPRICIGMLAMMEKKMLTHILKKYTFE 488

 RESULT 14
 ID YRYA_CAEEL STANDARD; PRT; 518 AA.
 AC Q27519;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DB 468 RLAMEMKMLTILNKTYFE 488

RESULT 15
C912_ARATH STANDARD: PRT; 500 AA.
ID C912_ARATH STANDARD: PRT; 500 AA.
AC 065790; Q95Z03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 91A2 (EC 1.14.-.-).
GN CYP1A2 OR ATG637430 OR P6G17.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seedling;
RX MEDLINE=98281573; PubMed=9620263;
RA Mizutani M., Ward E., Ohta D.;
RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of
cytochromes P450,"
RT Plant Mol. Biol. 37:39-52(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Scheller C., Wandut R., Murphy G., Volckaert G.,
Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansgore W., Brandt P., Griwell L.A., Rieger M.,
Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
Kreiss M., Delseny M., Pulgomech P., Watson M., Schmidheini T.,
Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,
Langham S.A., McCullagh B., Billham L., Robben J.,
Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandebussche F.,
Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
Weltenegeger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
Holzer E., Brandt A., Peters S., Van Staveren M., Dirke W.,
Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koelter P.,
Berneiser S., Hempel S., Feldpausch M., Lambers S., Van den Daele H.,
De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
Pettett A., Rajandream M.A., Lyne M., Benes V., Reichmann S.,
Borova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fairman B., Granderath K., Dauner D., Herzl A.,
Neumann S., Agirito A., Vitale D., Liguori R., Piravandi E.,
Massenet O., Quigley F., Clabaud G., Muehlen A., Felber R.,
Schubel S., Hiller R., Schmidt W., Leharay A., Audouin S.,
Cheddi F., Cooke R., Berger C., Montfort A., Casacuberta E.,
Gibbons T., Weber N., Vandebol M., Barges M., Terol J., Torres A.,
Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bialke C.,
Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
Parnell L., Dedha N., Gnoj L., Schütz K., Huang E., Spiegel L.,
Serkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
Stonking T., Kallack J., Graves T., Harmon G., Edwards J.,
RA Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
Min P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hiller L.,
Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton K., Layman D.,
Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lohli M., Johnson A.,
Chen E., Maria M., Martienssen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";
RL Nature 402:769-777(1999).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78607; BAA28539.1; -
DR EMBL: AL035601; CAB38210.1; -
DR Interp: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 438 438 HEME (BY SIMILARITY).
FT CONFLICT 106 106 A -> V (IN REF. 1).
FT CONFLICT 127 127 I -> M (IN REF. 1).
FT CONFLICT 140 140 N -> I (IN REF. 1).
FT CONFLICT 454 454 S -> T (IN REF. 1).
SQ SEQUENCE 500 AA; 57555 MW; 0FBA53D207DEA2BA CRC64;
Query Match 11.8%; Score 316; DB 1; Length 500;
Best Local Similarity 25.2%; Pred. No. 5,4e-12;
Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;
RT 9 LFLPLLLPLSLLLPLILKLR--RNKTRFENLPKSGMPPLGCTIGYKPTATYGLD 66
DB 1 MLTYFLTP-----LFLVLYSKFLYKSTQRFNLPSPSPSPFGH-LHMKP---PIHR 50
QY FMOOHVSKYKIRYSNLFGEPTV-----SADAGLRFITONGRLEFECSPRSIG 117
DB 51 LQGRYNGQGPRI-SLRFSGRRVYVITSPSLAGSEFTGONDYLSRPLDTAKYVAYNH 109
QY 118 GLIGKSMVLVVDMDHDMST-SLNFSLHARLRTI--LKDVRRHTLFLYDSMOONSIF 174
DB 110 TVYGV---PAPYGDHNRNLRICSOELTSSHRILNFOHINKDELIRLRLSRYTQS-- 163
QY 175 SAODEAKKPLF-----NLMAKHITMSMDPG-----EETFOLEKKEVYTFKGV 216
DB 164 ---NESNDTHILEPLSLDLTNNIVRYATGKRRYGGDVNNKEEALFK----- 211
QY 217 VSAPLNLPGTAYRKALQSRAT-----ILKFERKMEERKLDIKEEDQEEFEKTEDEA 269
DB 212 -----LVYDIAMVSGANHSADYLPILKLGKMFKEKVKRAG-----KSMDDI 253
QY 270 ESKSDPKVKQRDDDLGLVNLKHSNLTQILD-----LISLPLRAGHETSSVAYALAI 324
DB 254 LQRLDLDECRRDKGNTVMNHLISLQOOQPEPYTVIILKIGMSMMLAGTTSVATLEWAM 313
QY 325 FFLQAPKVAWEELREELHETARAKKEGESEL-NMWDYKMDFTQCVINETLRGNVRF 383
DB 314 ANLRMPVELEKRSF-----IDKIGKRLIDESDIANVPLQNVSTFRLFPYAPF 367
QY 384 L-HRKALKDVRYKGYDIPSGKVLPIYISAVHLNDSRYDQNLNPMFMQOONGASSSGS 442
DB 368 LIPRSPDTDMKIGGYDVPRPIYVNMAMAIHRDPELWEEBEKFNPDY---NDGC----- 419
QY 443 GSSTSTGNNT-----MPFGGPRPLGAGSELAKEMAVFTHHVLVKRWELAEDDQ 492
DB 420 -----GSDYVYVKIMPFGNGRRTCPGAGIGQRIIVTLALGSLIQCEWENVKGBE 468

Search completed: June 10, 2003, 21:57:52
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:54:48 ; Search time 49 Seconds
(without alignments)
2157.187 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681

Sequence: 1 MFETEHHTLPLLLPSLLS.....FAFPVDFPNGLPIVSRLL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	10	064989 arabidopsis
2	2677	99.9	513	10	09SCQ9 arabidopsis
3	951	35.5	474	10	09LKH7 vigna radia
4	866	32.3	512	10	09FX29 arabidopsis
5	843.5	31.5	491	10	094IA6 arabidopsis
6	822	30.7	478	10	09LNT3 arabidopsis
7	813	30.4	382	10	09LY89 arabidopsis
8	813.5	30.3	490	10	094IWS arabidopsis
9	789	29.4	464	10	094IWS arabidopsis
10	756.5	28.2	481	10	08MOR4 arabidopsis
11	726	27.1	465	10	09FMA5 arabidopsis
12	712	26.6	465	10	094OV4 arabidopsis
13	711	26.5	465	10	09LH81 arabidopsis
14	672.5	25.1	467	10	0949P1 arabidopsis
15	658	24.5	463	10	09FH76 arabidopsis
16	632	23.6	457	10	065624 arabidopsis

17	631	23.5	518	10	09FT38 arabidopsis
18	631	23.5	735	10	09IG17 arabidopsis
19	622	23.2	443	10	09LJK2 arabidopsis
20	611	22.8	482	10	081077 arabidopsis
21	606	22.6	479	10	08VZC2 arabidopsis
22	597.5	22.3	485	10	08W4T9 taxus cuspi
23	596.5	22.2	497	10	09AXM6 taxus cuspi
24	589	22.0	486	10	09L169 arabidopsis
25	581	21.7	487	10	023384 arabidopsis
26	581	21.7	496	10	09FOV4 arabidopsis
27	579.5	21.6	474	10	09SHY7 arabidopsis
28	579.5	21.6	485	10	09SJH2 arabidopsis
29	577	21.5	455	10	09LXH8 arabidopsis
30	576.5	21.5	490	10	09CSY3 arabidopsis
31	575.5	21.5	464	10	004949 arabidopsis
32	569.5	21.2	477	10	09LVT7 arabidopsis
33	547.5	20.4	489	10	09ZV72 arabidopsis
34	546.5	20.4	444	10	08S917 arabidopsis
35	546.5	20.4	489	10	09CSY2 arabidopsis
36	518	19.3	504	10	09SNG3 arabidopsis
37	502	18.7	499	10	09AXH9 arabidopsis
38	497.5	18.6	460	10	09STN2 arabidopsis
39	453	16.9	497	11	08VIL0 ratulus norv
40	435	16.2	224	10	09SDM6 helianthus
41	421	15.7	394	10	08W1S6 zea mays su
42	414	15.4	394	10	08W1T1 zea mays su
43	414	15.4	394	10	08W1S9 zea mays su
44	413	15.4	394	10	08WVJ5 zea mays su
45	412	15.4	394	10	08W1V5 zea mays su

ALIGNMENTS

RESULT 1

ID 064989 PRELIMINARY; PRT; 513 AA.

AC 064989:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Steroid 22-alpha-hydroxylase.
GN DWFA.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS-2;
RX MEDLINE=98158690; PubMed=9490746;
RA Cho S., Dalkes B.P., Fujioaka S., Takatsuto S., Sakurai A.,
RA Feldmann K.A.;
RT "The DWFA gene of Arabidopsis encodes a cytochrome P450 that mediates
RT multiple 22alpha-hydroxylation steps in brassinosteroid
RT biosynthesis."
RL Plant Cell 10:231-243(1998).
RC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF044216; AAC05093.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA: 58867 MW; B1639BD9A4DA6F3 CRC64;

Query Match 100.0%; Score 2681; DB 10; Length 513;

Best Local Similarity 100.0%; Pred. No. 1.6e-184;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFETEHHTLPLLLPSLLSLLFLILKRRNRKTRFNLPFGKSGWPFLETIGYLRKY 60
DB 1 MFETEHHTLPLLLPSLLSLLFLILKRRNRKTRFNLPFGKSGWPFLETIGYLRKY 60

QY 61 ATTGADPMOQHVSKYKTIYNSNLFGPEPTIVSADAGLNFIQNEGRLEFCSPRSIGIL 120
 DB 61 ATTGADPMOQHVSKYKTIYNSNLFGPEPTIVSADAGLNFIQNEGRLEFCSPRSIGIL 120
 QY 121 GKMSMLVVGDMHMRMSISLNFLSHARLRTILKDVRRHRLFVLDMSQOQNSIFSADQEA 180
 DB 121 GKMSMLVVGDMHMRMSISLNFLSHARLRTILKDVRRHRLFVLDMSQOQNSIFSADQEA 180
 QY 181 KKFENLMAKHIMSMDGPEETEOLKKEVYFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 DB 181 KKFENLMAKHIMSMDGPEETEOLKKEVYFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 QY 241 FIERKMEERKIDKEEDQEEVEVTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQ 300
 DB 241 FIERKMEERKIDKEEDQEEVEVTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQ 300
 QY 301 ILDLILSLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKEGSELNMD 360
 DB 301 ILDLILSLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKEGSELNMD 360
 QY 361 YKKMDFQCVINETRLGNVVRFLHRRKALKDVRKYGDIPSGMKVLPVISAHLNDSRYD 420
 DB 361 YKKMDFQCVINETRLGNVVRFLHRRKALKDVRKYGDIPSGMKVLPVISAHLNDSRYD 420
 QY 421 QPNLFNPMWQOQNGASSSGSSTGNNYMPFGGGRPLCAGSELAKLEMAVFIHILV 480
 DB 421 QPNLFNPMWQOQNGASSSGSSTGNNYMPFGGGRPLCAGSELAKLEMAVFIHILV 480
 QY 481 LKFNMELAEDDQPPAFPFVDPNGPLIRVSRIL 513
 DB 481 LKFNMELAEDDQPPAFPFVDPNGPLIRVSRIL 513

RESULT 2

Q9SC09 PRELIMINARY: PRT: 513 AA.
 AC Q9SC09
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Steroid 22-alpha-hydroxylase (DMF4) (At3g50660/T3A5_40).
 GN T3A5_40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bloeker H., Meves H.W., Lemcke K., Mayer K.F.X., Quettler F.,
 RA Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamliya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamanura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

RA Hayashizaki Y., Ishida J., Jones T., Kamliya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamanura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL132979; CAB62435.1; -
 DR EMBL: AF412114; AAL06567.1; -
 DR EMBL: AY090266; AAL90927.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Cytochrome; Oxidoreductase.
 SQ SEQUENCE 513 AA: 58867 MW: 61639BDD9A5D7C93 CRC64;

Query Match 99.9%; Score 2677; DB 10; Length 513;
 Best Local Similarity 99.8%; Pred. No. 3e-184;
 Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEETHTHTLLPLLLPILSLILFLILKRRNRKTRFNLPPKSGMPFGITGYLKYPT 60
 DB 1 MEETHTHTLLPLLLPILSLILFLILKRRNRKTRFNLPPKSGMPFGITGYLKYPT 60
 QY 61 ATTGADPMOQHVSKYKTIYNSNLFGPEPTIVSADAGLNFIQNEGRLEFCSPRSIGIL 120
 DB 61 ATTGADPMOQHVSKYKTIYNSNLFGPEPTIVSADAGLNFIQNEGRLEFCSPRSIGIL 120
 QY 121 GKMSMLVVGDMHMRMSISLNFLSHARLRTILKDVRRHRLFVLDMSQOQNSIFSADQEA 180
 DB 121 GKMSMLVVGDMHMRMSISLNFLSHARLRTILKDVRRHRLFVLDMSQOQNSIFSADQEA 180
 QY 181 KKFENLMAKHIMSMDGPEETEOLKKEVYFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 DB 181 KKFENLMAKHIMSMDGPEETEOLKKEVYFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 QY 241 FIERKMEERKIDKEEDQEEVEVTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQ 300
 DB 241 FIERKMEERKIDKEEDQEEVEVTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQ 300
 QY 301 ILDLILSLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKEGSELNMD 360
 DB 301 ILDLILSLFAGHETSSVAIALAIFFLQACPAVEELREHELEIARAKKEGSELNMD 360
 QY 361 YKKMDFQCVINETRLGNVVRFLHRRKALKDVRKYGDIPSGMKVLPVISAHLNDSRYD 420
 DB 361 YKKMDFQCVINETRLGNVVRFLHRRKALKDVRKYGDIPSGMKVLPVISAHLNDSRYD 420
 QY 421 QPNLFNPMWQOQNGASSSGSSTGNNYMPFGGGRPLCAGSELAKLEMAVFIHILV 480
 DB 421 QPNLFNPMWQOQNGASSSGSSTGNNYMPFGGGRPLCAGSELAKLEMAVFIHILV 480
 QY 481 LKFNMELAEDDQPPAFPFVDPNGPLIRVSRIL 513
 DB 481 LKFNMELAEDDQPPAFPFVDPNGPLIRVSRIL 513

RESULT 3

Q9LKH7 PRELIMINARY: PRT: 474 AA.
 AC Q9LKH7
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Cytochrome P450.
 GN CYP90A2.
 OS Vigna radiata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.

RT "Genomic sequence for Arabidopsis thaliana BAC T12C24 from chromosome
RT I.",
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AC025417; AAF8087.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450_2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 478 AA; 55058 MW; F67A5B9F2FD28558 CRC64;

Query Match 30.7%; Score 822; DB 10; Length 478;
Best Local Similarity 35.6%; Pred. No. 5,6e-51;
Matches 181; Conservative 95; Mismatches 187; Indels 46; Gaps 8;

QY 12 LLLPLSLSLFLFLILKRRNRKRTFNLPGKSGMPLFEGTIGYGLKPYATATLDFNQOH 71
DB 4 LLIVWSILLISITIMVYSWNRKPKRGKLPQSGMKFPLGSGIQDFKRNKTSIDIPFKER 63
QY 72 VSKYGIKYSRLNLEGEPTVTSADAGLNPLFIONEGRLFECSPNSIGGILGKMSVLVGD 131
DB 64 VKKGIPIFKTNLVRPIVSTADLSYFVFNQGRCSQSWYPDFTFIQKKNVSLHGF 123
QY 132 MHRMRISLINFSLHALRITLTKLDVERHILFVLDSMQQNSIFSAODEAKFTTNLMKH 191
DB 124 MYKYLKNNVLTLLFGHDLKRLMPQVEMTANKRLLELMSNODSVELKDATASMTFLYAK 182
QY 192 IMSADPEELETQKKEVTFPMKGVNSAPLNLPTAYAKALQSRAITLKFIERKMEERKL 251
DB 183 LISHD- DKSENNIRANFVAFIGSLIFRPDIPETAHNKLCQAGAKMKIRNLDERR- 240
QY 252 DIKEEDDEEEVKTEDAEKMSDHYAKQRTD--DDLGLVYLNKSNLSTEOI-LDLILSL 308
DB 241 -----ENPKRNSDFPDVYIEIQKEGTLITEIALDLMLFVL 277
QY 309 LFAHEHSSVAIALAIFLQACPRAVELEBEHEIARAKKELGESLNDMDYKKMPTQ 368
DB 278 LFAFEETSLALTAIAFKLSDPFLVRLTEHEHTLR-MNEADSLGTWEKYSMTYTF 336
QY 369 CVINSETRLGVNVAFELRKALKDVRK-----GYDIPSGMKVLPVISAHLNDSRYDOP 422
DB 337 QFIETIRLANIVAFIRKALRDKKEFVNDYITIPAGNAVCPVAHNLNEAKKDP 396
QY 423 NLEPNWMOQONNAGSSGSGSFSTWNNVMPFGGPRLAGSELALENAVFTIHLVYLK 482
DB 397 LVFNPSNWE-----GSKVTNASKHFMFAGGGRFCGTGFTLQMAAFHSLVTK 446
QY 483 FNNELADDDQPAFPVDFPNGRLIRYSR 511
DB 447 YRWEIKNGNTTRTGPLOFPNGYHVKLKH 475

RESULT 7
Q9LY89 PRELIMINARY; PRT; 382 AA.
AC Q9LY89;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 43.9 kDa protein.
GN F18022_190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.,
RA Rud S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
NN [2]

RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL163817; CAB87779.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Hypothetical protein; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 382 AA; 43889 MW; 1BC5685AB73BAE30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;
 Best Local Similarity 34.2%; Pred. No. 1.3e-50;
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

```

QY 45 GMPFGETIGYLPKPTATITGDFPMQOHSYKRTIRSNLFGPTVSDAGINRITONE 104
DB 2 GMPFGETISFPRKHSRDSIGTFLOQVSRGKVKRSNIGCGKAVVSCDOEIMFTIQNE 61
QY 105 GRLFCSTPRSTIGILGKMSMLVLDGMDRMSISLNLFLSHARLTILKDVERTLTFY 164
DB 62 GRLFTSDYPRKAMHMDILGKYSLLATGELIHRKLNKNTISFINLTSKRPFLCAENLSTI 121
QY 165 LDMQOANSIFSAODEAKKFTFNIAKHTMSMDPGEETEQLKKEYVTFMGVSAPIALP 224
DB 122 LKSWNCRVEVEHKEVKLFTLSVMYNQLSTKPEDPARLYVLODFLSYMKGFISLPIPLP 181
QY 225 STAYHKAQSRTATILKTERKMEERKLDIKEEDQEEVEKTEDEAKMSKSDHVKORTD 284
DB 182 GTGTNAIKVRSNNIHNQNAIIEEDMNNAIREDLDLSTISMED----- 224
QY 285 DLGQVNLKHSNLSLSTEQILDLISLFLAGHETSSVAIALAFLQACPKAVEELREHLEI 344
DB 225 -----EEHAAI 230
QY 345 ARAKKEGSELNMDYRKMDPTOCYINETLRGNNVFLHRAKLDVRYKGDIPBSGM 404
DB 231 -RAKKGDGEL-LNEDYQKKEFGCVISELRGNNYKTYHRAKTHDINKNEYVIRPGMK 288
QY 405 VLPYISAVHLDNSYDDPNIENFRKQOONNGASSGSGSFSTGNNVMPFGGPRCLAG 464
DB 289 VFPIETAVHLDPPLSHNEPFEFENPRMKTTP-----AFGGVAVCVRG 329
QY 465 SELAKLEMAVFIHHLVLFKNMELAEEDQPPAPFVDPENGLPIRV 509
DB 330 GEIGKLIQIAFFLHLVLSYRMKIKSDMPIAHPIYVEKRMGLEI 374

```

RESULT 8

ID 0941W5 PRELIMINARY; PRT; 490 AA.
 AC 0941W5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 GN P0419B01.5.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:p0419B01.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AP003244; BAB56089.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.

DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 490 AA; 54824 MW; 9BC2853BBAFBB88P CRC64;

Query Match 30.3%; Score 813.5; DB 10; Length 490;
 Best Local Similarity 34.2%; Pred. No. 2.4e-50;
 Matches 173; Conservative 106; Mismatches 136; Indels 71; Gaps 9;

```

QY 27 LTKRRNR-----KTFENLPKSGKMPFGETIGYLPKPTATITGDFPMQOHSYK 77
DB 28 LTKRRNRKAGSGKDAAAAMALPGSGFMPVGETLFEVSOASPREAVDRKRIHGS 87
QY 78 -IYSNLFGEPTIYASDAGINRFTLOMEGRFLFECSTYRSIGILGKMSMLVLDGMDRM 136
DB 88 AVFSSHLPFGATVTTAAAEVSRFLVQSDARAFVFWYRSLTELMGSSILLINGALQRRV 147
QY 137 RSISLNLNLSHARLTILKDVERTLTFVLDMSQOANSIFSAODEAKKFTFNIAKHTMSMD 196
DB 148 HGLVGAFFKSSHLKSQTLADMRRRLSPALSPFDSLSLHQLAKSVFEILVGLGLE 207
QY 197 PGEETEQLKKEYVTFMGVSAPIALPGRATYHKAQSRTATILKTERKMEERK----- 250
DB 208 AG-DEMQLKQOEFTVGLMSLPKLPGRITRSLQAKKARLQRTIIRERARRAAA 266
QY 251 -----LDIKEEDQEEVEKTEDEAKMSKSDHVKORTDDLGVLKHSNLSLSTEQI 304
DB 267 SPFRDAIDVILIGDSDE-----LTDELISDN 292
QY 305 ILSLFAGHETSSVAIALAFLQACPKAVEELREHLEIARAKKEGSELNMDYKMK 364
DB 293 MIDLMIPAEDEVVLTITLAVKFLSECPDLHLHOLEENIOLKRRKTDMEGT-LQMDYMSL 351
QY 365 DFTOCYINETLRGNNVFLHRAKLDVRYKGDIPBSGMVPLVSAVHLNSRYDDPRL 424
DB 352 SFYOHVITETLRGNIIGIMRAKAVDVEYKHILPKGCYVYRSHLDOTLDEPK 411
QY 425 FNPWRQOONNGASSGSGSFSTGNNVMPFGGPRCLAGSELAKLEMAVFIHHLVLFKN 484
DB 412 FNPWRKEND-----MSNSST-----PFGGGRQLCPGLDLARLKSIFLHLVTSFR 459
QY 485 WELAEEDQPPAPFVDPENGLPIRV 510
DB 460 W-VAEDHIVNPPYRLKRMPIRV 484

```

RESULT 9

ID 091IC5 PRELIMINARY; PRT; 464 AA.
 AC 091IC5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE-20363099; Pubmed-10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL: AF001307; BAB01922.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 464 AA; 52864 MW; A65E094665E51908 CRC64;
 SQ
 Query Match 29.4%; Score 789; DB 10; Length 464;
 Best Local Similarity 36.5%; Pred. No. 1.3e-48;
 Matches 163; Conservative 101; Mismatches 145; Indels 38; Gaps 8;

QY 40 PPKSGMPFGEITGYLKPATTLTGDFMOQHVSKGKRYNSNLGFEPTIYSADAGLRF 99
 DB 52 PHGSLGMPVIGETIEFVSAYSDRSPMDKRLMYGVFKSHLFGTATITSTDAEVRA 111
 QY 100 ILQNEGKLFECSPRISGIGILGKWSMLVGDHMRMSISLNLSHAARLTILKDYER 159
 DB 112 VLQSDSTAFVPEFYKTYRELMGKSSILLINGSLRRFHLGVSFLKSPILKAQIVRDMK 171
 QY 160 HTLFVLDSMOONSIFSAODEAKKFTFNLMKHMISMDGEEETQLKKEYTFMKGVYA 219
 DB 172 FLSESMDSDDQVLLQDVSKTYAFVLAALISVEKG-RDLELKEFEFNISGLMSL 230
 QY 220 PLNLPGTAYHKAQSRATILKFERKKEERKLDIKEEDQEEVEKTEDEAM---SKSDH 276
 DB 231 PINPFGQLHRSIQAKKNMVKQVERITEGRK--KTNKKEEDVIADQVYVILKDSSEH 288
 QY 277 VRKQRTDLDLGVILKHSNSTEOILDLILSLFAGHETSSVAIALAIFLOAPKAAVEE 336
 DB 289 -----LTFH-NLIANNMIDMI-----PGHDSVPVLLITLAKVFSDSPALNL 329
 QY 337 LREHELIARAKKELGSELSLMDYKHMDFQCYINETLRGNYVRLHRAKLDVRYKG 396
 DB 330 LTERNMKL-KSLKELTEPEPLYNDYLSLPTQKVTTELTRMGANYITGMRAMKQVEIKG 388
 QY 397 YDIPSGKVLVPIVSAVHLDSRYDQPNLFNFRMQOONNGASSSGSFSFTWGNVMPFG 456
 DB 389 YVLPKMGCPFLAYLSVHLDKLYESPYKFNFMWQERDMNTSS-----FSPFG 436
 QY 457 GGRLCAGSELAKLEMAVFIHHLVLFK 483
 DB 437 GGORLCPGLDLARLETSVFLHHLVTRF 463

RESULT 10
 OBMORA
 ID OBMORA PRELIMINARY; PRT; 481 AA.
 AC OBMORA;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative cytochrome P450-like protein.
 GN SB32H7.4.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Sorghum.
 OX NCBI_Taxid=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTX623;
 RA Park Y.-J., Ramakrishna W., Sanmiquel P., Emberton J., Bennettzen J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTX623;
 RA Liaca V., Young S., Koychok S., Messing J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF466301; AAL73972.1; -
 DR InterPro: IPR001128; Cytochrome_P450.

DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR SEQUENCE 481 AA; 54984 MW; 6145CFEFCOFD74FE CRC64;
 SQ
 Query Match 28.2%; Score 756.5; DB 10; Length 481;
 Best Local Similarity 33.5%; Pred. No. 2.9e-46;
 Matches 170; Conservative 104; Mismatches 190; Indels 43; Gaps 7;

QY 15 LPSSLILFLILK-----RRNRTRNLPPKSGMPFLGTYLKPATTLTG 65
 DB 6 LAALSVTLGALLLMAFMKMMNYGRGCEGMLLPKSGMLPFLGTYLFPASPLELV 65
 QY 66 DFMQHVSKGKRYNSNLGFEPTIYSADAGLRFILQNGRLPECSPPSIGILGKWSM 125
 DB 66 PEFKRRLFEFGPIFTNIVGEMIVSLDELARVLOQDERGFOIYPPSFMILLGADNM 125
 QY 126 LVLYGDMHRDMSISLNLSHAARLTILKDYERHTLFVLDSMOONSIFSADQAKKFTF 185
 DB 126 VSMGLPRLHRIINLRLFGPEALRLVLRDVORSARDELRSMLDRPEVEKRYSRMTL 185
 QY 186 NLMAKHMSMDGEEETQLKKEYTFMKGVYAPNLPGTAYHKAQSRATILKFERK 245
 DB 186 GYTAKKLISHD-DVAGSGSLMKCFDAMTGMKLSFICVGTAFYRCMOGRKNMVKLKO 244
 QY 246 MEERKLDIKEEDQEEVEKTEDEAKSKSDHVRKQRTDLDLGVILKHSNSTEOI-LDI 304
 DB 245 LDERR-----NGAERKTVDFDL-----VIDELDKPMSIMSESIALNL 282
 QY 305 ILSLFAGHETSSVAIALAIFLOAPKAAVEELREHELIARAKKELGSELSLMDYKYM 364
 DB 283 LFLLFASHETTSMLGTYLILKFLTONPSLOLTFEHEKIMERRAD-PSDITWEYKSM 341
 QY 365 DETQCVINETLRGNYVRLHRAKLDVRYKGYDIPSGKVLVPIVSAVHLDSRYDQPNL 424
 DB 342 KFTSHVHESLRLANLAPVYEQANDVHIKTTIPESKIMICPSAALNSVYEDPLA 401
 QY 425 FNPWMQOONNGASSSGSFSFTWGNVMPFGGPRLCAGSELAKLEMAVFIHHLVLFK 484
 DB 402 FNPWMKQTPPEVGS-----KDFMAFGGLRLCGAEGFAKLQAMFLHYLTNTR 452
 QY 485 WELAEDDQFAFPFDPFGRLPIRYSR 511
 DB 453 WKALSKGTMTLPGRLPFGFHTQLHK 479

RESULT 11
 O9FMA5
 ID O9FMA5 PRELIMINARY; PRT; 465 AA.
 AC O9FMA5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cytochrome P450 (Brassinosteroid-6-oxidase).
 GN BR60XI.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustos II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,450,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=21295570; PubMed=11402205;

RA Shimada Y., Fujioaka S., Miyauchi N., Kushihiro M., Takatsuto S.,
 RA Nomura T., Yokota T., Kamiya Y., Bishop G., Yoshida S.,
 RT "Brassinosteroid-6-oxidases from Arabidopsis and tomato catalyze
 RT multiple C-6 oxidations in brassinosteroid biosynthesis.";
 RL Plant Physiol. 126:770-779 (2001).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB009048; BAB08653.1; -
 DR EMBL: AB033868; BAB0858.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 SO SEQUENCE 465 AA; 53767 MW; 0C00459C9C866D1F CRC64;

Query Match 27.1%; Score 726; DB 10; Length 465;
 Best Local Similarity 32.3%; Pred. No. 4.3e-44;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

12 LLLPSSLLSLFLILKRRN--KTRRNLPKSGKWPFLGTTGYLKPATATLGDPMQ 69
 11 LLIIVSLCSALL-----RWMQMRKTKNGLPPTGMPGIFGTEFLKQGP-----NFMK 59
 70 OHVSKYKIRYSNLFGEPTIVSADAGLNRFILONGRLEFCSYPRISIGLIGKSMVLV 129
 60 NORLRYGSFRRSHLIGCPITLSMOSEVAKYLKSKGLVPGYQSMIDILGTGNMAVH 119
 130 GDMHDMRSISLNFSLHARLRTLLIKDVERHTLVDSMOONSIFSAODEAKKFTF---- 185
 120 GSHHLMKGSLLSLISSTMMDHLIPKVDHFRSYLDQMNLEVIDIDOKRHAFSL 179
 186 -----NLMAKHMSMDPEEETEELKKEVYTFMKGYSAPLNLFGLYHKALOSRAVILK 240
 180 TOIAGNLRKPFV-----EEFKTAFKLVGLTSPDIDPGTYRCGIOARNNIDR 229
 241 FIEKMERKIDIEEQEVEEYKTEDEAEMSKSDHYRKORTDDLLGWLKHSN-----L 236
 230 LIRLMEKERR-----DSGE-----TTDMGLYLMKKEGNNRPL 262
 297 STEQILDLILSLFAGHETSSVAIALAIFLOCPKAVEELREHETIARAKKEISEL 356
 263 TDEIRDOQVYTLISGEVYVSTSMALKYLHDHPNALQRLRAEHLAFREKRQ--DEPL 320
 357 NMDYKMDTQCVINETTLRGVNRFLHKKALDKVRYKGYDIPSGKVLPIVSAVHLN 416
 321 GLEDVSMKFTRAVIYETSLRATVNGVLRKTRDLEINGYLPRKRIYVITREINYDA 380
 417 SRYDQPLFPMRMOONNGASSSGSSTWGNMMPGGRPCAGSELAKLENAVPI 476
 381 NLEDPLIFPMRMAKKSLESO-----NSCFVGGGTGLCPGKELGIVEISSFL 429
 477 HHLYLKFNNELADDDPFAFPVDFPGLPIRVS 510
 430 HFFVTRYRWEELIGDELWFPFVFAKGFHLKIS 463

RESULT 12

O940V4

AC 0940V4; PRELIMINARY; PRT; 465 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE AT3930180/T20E20.6
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawat J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Saitou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawat J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Saitou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY052655; AAK96559.1; -
 DR EMBL: AY063728; AAL36078.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 SO SEQUENCE 465 AA; 53814 MW; 280A21D0712FCAL1 CRC64;

Query Match 26.6%; Score 712; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 4.3e-43;
 Matches 163; Conservative 94; Mismatches 194; Indels 56; Gaps 8;

14 LPLSLSLFLIL-----KRNKRTFNNLPKSGKWPFLGTTGYLKPATATLGD 67
 3 IMMILGLVITICTALLRWNRKSKGLPPTGMPGIFGTEFLKQGP-----DE 57
 68 MOOHVSKYKIRYSNLFGEPTIVSADAGLNRFILONGRLEFCSYPRISIGLIGKSMVLV 127
 58 MNORLRYGSFRRSHLIGCPITVSDPAENRITLNNESGLVAGYQSMIDILGTGNIAA 117
 128 LVGDHMRDRSISLNFSLHARLRTLLIKDVERHTLVDSMOONSIFSAODEAKKFTNL 187
 118 VGPSHRLMRGSLISLISPTMKDHLIPKIDPMNLYLCGMDLETVIOEKTHMA-L 176
 188 MAKHMSMDPEEETEELKKEVYTFMKGYSAPLNLFGLYHKALOSRAVILKTERKME 247
 177 SSLDIAETLKKPEVEEYRTEFEKLVGLTSPDIDPGTYRCGIOARNNIDRLTELMQ 236
 248 EKKIDKEDEEVEEYKTEDEAEMSKSDHYRKORTDDLLGWLKHSN-----LSTEQILD 303
 237 ERK-----ESGETF-----DMLGYLMKKEGNNRPLLDKTERD 269
 304 LILSLFAGHETSSVAIALAIFLOCPKAVEELREHETIARAKKEISELMDYK 363
 270 QVITLISGEVYVSTSMALKYLHDHPKALEERREHILAIERKR--DEPLTDIKS 327
 364 MDTQCVINETTLRGVNRFLHKKALDKVRYKGYDIPSGKVLPIVSAVHLNDRYDQPN 423
 328 MKFTRAVIYETSLRATVNGVLRKTRDLEINGYLPRKRIYVITREINYDLSYEDPM 387
 424 LFNWRMOONNGASSSGSSTWGNMMPGGRPCAGSELAKLENAVPIHHLYLKF 483
 388 IFNWRMERKSLSKS-----YFLFGVGYRLCGKELGISEVSEFLHYVYTK 436
 484 NNELADDDPFAFPVDFPGLPIRVS 510
 437 RWEENGEDKLVMPFVFAKGFHLKIS 463

RESULT 13

O9LH81

AC 09LH81; PRELIMINARY; PRT; 465 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cytochrome P450.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA MEDLINE=20363099; PubMed=10907853;
 RX Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AP002060; BAB02270.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KM Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53862 MW; 630A21D0765EED0D CRC64;
 Query Match 26.5%; Score 711; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 5.1e-43;
 Matches 163; Conservative 93; Mismatches 195; Indels 56; Gaps 8;
 QY 14 LLLPSLLSLFLLL-----KRRNRKTRFNLPPGKSGMPFLGTYGLKPYTATTLGDF 67
 DB 3 IMMMLGLVITVLCCTALRNWNRKSKLPPGTWMPFLGTEFLKQGP-----DF 57
 QY 68 MOCHVSKYKIRSNLGEPTIVSADAGLNRFILQNEGRLEECSPRSIGILGKWSMLV 127
 DB 58 MKNQRLRYGSEFFKSHIIGCPITVGMDELNRYIILNNEKGLVAGYPOGMDLITGICNIAA 117
 QY 128 LVGMHMDMSISLNFISHARLRTLLKQVERHTLEVLDSQOONSIFSAODEAKKFTNL 187
 DB 118 VHGSHRLMRGSLSLISPTMMKDLPLPKIDFMKNRYLCGWDDELTVDIOEKTKHMAF-L 176
 QY 188 MAKHMSMDPGEETEDOLKKEYVTFMKGVSAPLNPGTAYHKAQSRATILKFERKME 247
 DB 177 SSLQIAETLKKPEVEYRTFFKLVGTLSVPIDIPGTNYSRGFQANNIDRLITELMO 236
 QY 248 ERKLDIKEEDOEVEEVTEDEAEMSKSDHVKORTDDLLGWLKHSN-----LSTEQITLD 303
 DB 237 ERK-----ESGETFT-----DMLGYLKKKEENRLLTDKEIRD 269
 QY 304 LILSLLAGHTSSVAIALAIFFLQACPKAVEELREELHETARAKKELGESLMDYDK 363
 DB 270 QVATILSGYETVSTSMALKYLHDHPKALELRREILAIRKRKP--DEPLTDDIKS 327
 QY 364 MDTQCYINETLRGNVYRFLHRAKALQVRYKGYDIPSGWVLPYISAVHIDNSYDDPN 423
 DB 328 MKPLRAVIFETSRLATIYNGVLRKTHDLNGLILPGRWRIYYTRREINIDTSLYEDPM 387
 QY 424 LFNFMWQOONNGASSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHLLVKE 483
 DB 388 IFNFMWMEKSLSEKS-----YFLFGGVRRCPEKELGISVSSFLHFTYKY 436
 QY 484 NMLAEDDQPPAFVDFPNGLPYRS 510
 DB 437 RMEENGEDKLWFPFVSPAPKGYHLKCS 463

RESULT 14
 ID Q949P1 PRELIMINARY; PRT; 467 AA.
 AC Q949P1;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative cytochrome P450 protein.
 GN AT4G19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carlincl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Setou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full Length cDNA of gene At4g19230 (GI:7268718)."
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carlincl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Setou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY050980; AAK93657.1;
 DR EMBL: AY091446; AAM14385.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KM Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 467 AA; 53037 MW; 2F4230446536D955 CRC64;
 Query Match 25.1%; Score 672.5; DB 10; Length 467;
 Best Local Similarity 32.3%; Pred. No. 3e-40;
 Matches 164; Conservative 96; Mismatches 183; Indels 65; Gaps 15;
 QY 13 LLLPSLLSLFLLLKRRNRKTRFNLPPGKSGMPFLGTYGLKPYTATTLGDFMOOHV 72
 DB 10 LFAAGSLFLFLKCLISORFFGSKLPLPGTWMFVGEF---FOLYSQDP-NVFRSQK 65
 QY 73 SKYGRKIRSNLGEPTIVSADAGLNRFILQNEGRLEECSPRSIGILGKWSMLVLDGM 132
 DB 66 KRYGSVFKNHVLGCPVMTSSPPAAKFEVLVTSHLKPPFPAPKERMKGQAIFPQGD 125
 QY 133 HRDMRSISLNFISHARLRTLLKQVERHTLEVLDSQOONSIFSAODEAKKFTNLMAKH 192
 DB 126 HAKLKLIVRAAMPESIRN-NVPTDISIADQSLRSR-EGTMINTYDEMKITYTNVALIST 183
 QY 193 MSMDPGEETEDOLKKEYVTFMKGVSAPLNPGTAYHKAQSRATILKFERKMEER 249
 DB 184 F-----GKDELVLREDLKRCYIYLEKGYNSMPVNLPETLHKSKAKKRELSQLARLSER 239
 QY 230 KLDIKEEDOEVEEVTEDEAEMSKSDHVKORTDDLLG-VWLKHSNLSLTDLLDILSL 308
 DB 240 R-----QNGSSH-----NDLLSGFMDKDELTDEQIADNITGV 272
 QY 309 LFAGHETSSVAIALAIFFLQACPKAVEELREELHETARAKKELGESLMDYDKKDFNQ 368

Db 273 IFARDTTASVSMILKYIAENPNVLAETEOMAI-RKDEGES-LTWGDTKMPILTS 330
 QY 369 CVINETRLGNVVRFLHRKALKDVRKGYDIPSGWVLPYISAVHLNDSRYDQPNLFNFM 428
 Db 331 RVIOETLRVASILSTFPEREVEVEGYLIPRGWVLPFLFRNIHSADIENPKGFDS 390
 QY 429 RMOQNNNGASSSGSFSTWGNMYPFGGPRLCAGSELAKEMAVFIHHLVKFNMEL- 487
 Db 391 RE-----VAPKPMTEMPFGSGHSCPNELAKLEMSIMHILTKYSWSIV 437
 QY 488 -AEDD---OPAFPFVDFPNGLPPIRYSR 511
 Db 438 GASDGIQYGFALP-----QNGLPILYAR 461

RESULT 15

Q9FH76 PRELIMINARY: PRT: 463 AA.
 ID 09FH76;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cytochrome P450 (AT5G45340/K9E15_12).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carlucci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kwal J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB020744; BAB10255.1;
 DR EMBL: AY065065; AAL57698.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450.1.
 DR PRINTS: PR00385; P450.1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 463 AA; 52366 MW; CCD17293F553P812 CRC64;

Query Match 24.5%; Score 658; DB 10; Length 463;
 Best Local Similarity 31.8%; Pred. No. 3.3e-39;
 Matches 163; Conservative 95; Mismatches 184; Indels 70; Gaps 15;

QY 14 LIPSLSLFLFLILK-----RRNKKTFENLPCKSGWPLGERTIGYILKPYTATLGDPM 68
 Db 6 LFLTLSSAALFLCLRLFIAGVRSSSTKRLPPTMGPIYVGET---FQLTSDDP-NVFF 61
 QY 69 QOHVSKYGIKIRSNLFGPTIVSADAGLNFTIIONEGRLFECSYPSRIGILGKMSMLVL 128
 Db 62 AAKRRRGSVKTHVLGCPGVMISSPPAKFVLVTKSHLKPFPFPASKERMILGQALFFH 121
 QY 129 VGMHMRDRISINFLSHARLFTLLKDVHRHTFLVDSWQNSIFSAQDEAKKFTFNLM 188

Db 122 QGDYSKRLKLVLRAMPDAIN-WPHIESIAOBSLNSWGOTQNTYQ-EKRYTFNVA 179
 QY 189 AKHISMDFEET---EOLKKEYTFMKGVVSAPLNPGRAYHKALOSRTILKFIERK 245
 Db 180 LISIL---GKDEYVYRDDLRKYVILKGYNSMFINLPGLFHKAKARELQIILANI 235
 QY 246 MEERLIDKEEDQEEVEKTEDEAMSKSDVRRKQRTDLDLGWVK-HSNLSTEQIDL 304
 Db 236 LSKRR-----QNSSHT-----DLGSMEDKMGILTDEQIADN 268
 QY 305 ILSLFAGHETSVAIALAIFELQACPAVEELREHLEIARAKKEGSELMNDYKRM 364
 Db 269 IIGVIFARDTTASVLMILKYITADNPVLEAVEEOMAIRKDE-GE-S-LTWEDTKM 326
 QY 365 DFQCVINETRLGNVVRFLHRKALKDVRKGYDIPSGWVLPYISAVHLNDSRYDQPNL 424
 Db 327 PLIVRVIOETLRVAITLSTFEREVEVEGYLIPRGWVLPFLFRNIHSADIENPKGF 386
 QY 425 RMOQNNNGASSSGSFSTWGNMYPFGGPRLCAGSELAKEMAVFIHHLVKFNMEL 484
 Db 387 FDPGRF-----VAPKPMTEMPFGSGHSCPNELAKLEISVLHILTKYR 433
 QY 485 WELAEDDQ-----PFAFPVDFPNGLPPIRYSR 511
 Db 434 WSIIVGSDGIQYGFALP-----QNGLPILYAR 461

Search completed: June 10, 2003, 21:58:53
 Job time : 52 secs

